

; Seq. Id. Nos. 3 and 11 from USSN 09/770,693 joined  
; in the order SID 11 followed by SID 3  
693\_11\_3 SEQ ID NO: 11 SEQ ID NO: 3  
MGFFLFSOMPSEFLVSTLLFLIISHSHAQNSRMSLNTSGLGASTMOISIGGAGGNGLGTSRONAGL  
GGNSALGLGGGNQNTVNOGLAGLITCMMMMSMMGGGLMGGGLGGLGNGLGGGGLGEGLSNALNDML  
GGSLNTLGSKGNNTSTNSPDLQALGINTSQDDSTSGTSDSDSDPQQLLKMFSEIMQSLFGDG  
QDGTQSSSGGKOPTGEQNAKKGVTDALSLGMLGNSLQLLGGGGLGGGAGNAGTGLDGSLSGKGLQ  
NLGGPDYQQLGNVAGTGTGMRAGTQALNDIGTHRSSTRSFYNKGDRAKKEIGQFMQYPEVFGKPOY  
OKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMTKRPWAGDTGNGNLQARGAGGSSSLGIDAM  
MAGDAINNNALGKLGAA1

Please check the joined  
Sequences above to make sure  
I got it right!  
Thanks,  
Thom

Point of Contact:  
Thomas G. Larson, Ph.D.  
703-308-7309  
CM1, Rm. 6 B 01

GenCore version 5.1.3

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using 'sw model

Run on: January 22, 2003, 16:26:07 ; Search time 35 Seconds

(without alignments)  
2572.645 Million cell updates/sec

Title: 693\_11\_3

Perfect score: 2248

Sequence: 1 MGFFLFQMPSEFLVSTLL.....DAMMAGDAINMALGKLGAA 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1181.5	52.6	382	2	Q9FCY8
2	1002.5	44.6	365	2	Q9KH45
3	840.5	37.4	356	2	Q93FY8
4	736.5	32.8	344	2	Q9EXP0
5	712	31.7	142	2	Q9LAW6
6	264.5	11.8	1489	16	O53559
7	255	11.3	591	16	O07224
8	254	11.3	622	16	Q8VKN3
9	251.5	11.2	628	16	Q8VJ19
10	250	11.1	635	16	Q8VK71
11	250	11.1	775	16	Q8VJ15
12	249.5	11.1	1217	16	Q8VJ19
13	249	11.1	731	16	O50415
14	248	11.0	1381	16	O53552
15	248	11.0	1384	16	Q8VIZ1
16	248	11.0	1715	16	Q8VIZ0

17	247.5	11.0	1079	16	O53557	053557 mycobacteri
18	247	11.0	588	16	O50396	050396 mycobacteri
19	246	10.9	1538	16	O53395	053395 mycobacteri
20	245.5	10.9	923	16	O53890	053890 mycobacteri
21	245.5	10.9	1665	16	O53215	053215 mycobacteri
22	244	10.9	1507	16	O8VJ23	O8VJ23 mycobacteri
23	241.5	10.7	879	16	O8VKD2	O8VKD2 mycobacteri
24	241	10.7	3659	16	Q98LN6	Q98LN6 rhizobium l
25	238.5	10.6	882	16	O53845	053845 mycobacteri
26	236.5	10.5	436	16	O8VIX6	O8VIX6 mycobacteri
27	236.5	10.5	714	16	O53556	053556 mycobacteri
28	236.5	10.5	853	16	O53439	053439 mycobacteri
29	235.5	10.5	840	5	Q95YF6	Q95YF6 patinopecte
30	235	10.5	439	16	O06287	O06287 mycobacteri
31	234.5	10.4	1408	16	O8VK17	O8VK17 mycobacteri
32	234	10.4	2174	16	O92U08	O92U08 rhizobium m
33	233.5	10.4	349	10	O9LSP2	O9LSP2 arabidopsis
34	233.5	10.4	615	16	O05806	O05806 mycobacteri
35	232	10.3	837	16	O53684	053684 mycobacteri
36	229.5	10.2	1329	16	O06810	O06810 mycobacteri
37	228.5	10.2	486	10	O9AR23	O9AR23 oryza sativ
38	228.5	10.2	667	16	O53415	053415 mycobacteri
39	228.5	10.2	904	5	O76271	O76271 mytilus edu
40	228	10.1	783	16	O53809	053809 mycobacteri
41	227	10.1	496	16	O53631	053631 mycobacteri
42	227	10.1	533	16	O8VKR5	O8VKR5 mycobacteri
43	226.5	10.1	491	16	O06818	O06818 mycobacteri
44	226.5	10.1	576	16	P71664	P71664 mycobacteri
45	226.5	10.1	650	16	O8VJW1	O8VJW1 mycobacteri

## ALIGNMENTS

## RESULT 1

Q9FCY8 PRELIMINARY: PRT; 382 AA.  
 AC Q9FCY8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE HrpN.  
 GN HrpN.  
 OS Erwinia stewartii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pantoea.  
 OX NCBI\_TaxID=66271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SS104;  
 RA Frederick R.D., Ahmad M., Majerczak D.R., Arroyo-Rodriguez A.S.,  
 RA Manulis S., Coplin D.L.;  
 RT "Genetic organization of the Pantoea stewartii subsp. stewartii hrp  
 RT gene cluster and sequence analysis of the hrpA, hrpC, hrpN and wlsE  
 RT operons.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF282857; AAC01466.1; ...  
 SQ SEQUENCE 382 AA; 37878 MW; 8BA8C723F2DA38BB CRC64;

Query Match 52.6%; Score 1181.5; DB 2; Length 382;  
 Best Local Similarity 62.1%; Pred. No. 1.2e-60;  
 Matches 251; Conservative 44; Mismatches 86; Indels 23; Gaps 11;  
 QY 35 MSNTSLGCASTMQISIGGAGNNGLLGTSRQNAAGLGSALGLG-GGNQNDVTYNQLAGL 93  
 Db 1 MSNTSLGTSALQVTL---GGNGLMGTLDTLDGLGLSQPGLGEGKGHNESIDLLAAA 57  
 QY 94 LTGMAMMSMMGGGLMGGLGGLGGLGGLGNLNDMLGGSINTLGSKGN 153  
 Db 58 LTGMAMMSMMGGGL---SSLLSGTGMGNPFGGSGAPGNTLSGTSG-----GSPGG- 109  
 QY 154 NTSTSTNPLDQALGNTNSTQNDSTSGTSTSDSSDPMOOLLKMFSEIMQSLFGDGDG 213

```

Db 110 -TTGAGSS-----LGLDPTQTGDDSLSGAGQSGMS-PMEOLMKIPADITOSLFGD-QDC 161
QY 214 TCGSSSGKQPTGEQNAKKYKGVTDALSLGMNGLSQLLNGSLGGGCGGAGTGLDGGSS 273
Db 162 ASGGNA-CRQPSDEONAYKKVTDALFAPMGGLSQVAGNGSEGLDGG-WGLG-CGNG 218
QY 274 LGGKGLQNLSPGVYQOGLNAVGTGIGMKAGIOALNDICTHRSSTRSPVNVKGDRAK 333
Db 219 LGGKGLQNLSPGVYQOGLNAVGTGIGMKAGIOALNDICTHRSSTRSPVNVKGDRAK 278
QY 334 IGFQMDQYPEVFGKPOYKQPGOEKVTDDKSWAKALSCKPDDDDGMPASMEQFNKAKGM 393
Db 279 VGFQMDQYPEVFGKPOYKQPNADSAVKTDKSWAEALSQPDGDDGMPASMEQFNKAKGI 338
QY 394 RPMAGDTGNGNLQARGAGSSGIGIDAMACDAINNMAKGLGAA 437
Db 339 SAMAGDNGNINLQARGAGSSGIGIDATLTCDAINNMAKGLGAA 382

RESULT 2
ID Q9KH45 PRELIMINARY; PRT: 365 AA.
AC Q9KH45;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Harpin.
GN HRPN.
OS Pantoea agglomerans pv. gypsophilae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=48984;
RN [1]
RP MEDLINE=21171042; PubMed=11277443;
RA Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;
RT "Genetic organization of the hrp gene cluster and dspAE/BF operon in
RL Erwinia herbicola pv. gypsophilae.";
RL Mol. Plant Microbe Interact. 14:431-436(2001).
DR EMBL: AF271716; AAF76342.1; -.
SQ SEQUENCE 365 AA; 36784 MW; 6F79B6BDDDD95AC2A CRC64;

Query Match 44.6%; Score 1002.5; DB 2; Length 365;
Best Local Similarity 57.9%; Pred. No. 2.3e-50;
Matches 223; Conservative 40; Mismatches 83; Indels 39; Gaps 14;

QY 35 MSLNTSGLCASTMQISIGGAGNGNLLGTSRQNALGNGSALGGLGGNNDTVNOLAGLL 94
Db 1 MSLNTPLGIPAMQISL-----GDSNGLPGPNLQNAVLN-----SMGDRSTDETINQALAG 52
QY 95 TGMNMMMSMMGGGLGGLGGGLGGLG--SCGLGCLSNALNDMLCGSLNT-LGSKG 151
Db 53 TAMMMNMMMSMMGGGLGGLGGLGGLG--SCGLGCLSNALNDMLCGSLNT-LGSKG 110
QY 152 GNNTTSTTNSPLDQALGINSNDSTSGTSTSDSS-----DPMQQLKMFSEIMQSL 206
Db 111 G-----AGSILDRALDL-----DPTADGCLSDSNGTYGMSPIDQLMKFAEVMQSM 157
QY 207 FGDGDDTQGGSSGKQPTGEQNAKKYKGVTDALSLGMNGLSQLLNGSLGGGCGGAGNAG 266
Db 158 FG-GGGDTSGRGT-CNQPTPDBONAVTKGVTDALFVAVMGGLSQGGS-GTGGHNGSIG 214
QY 267 TGLDSSGLGGKGLQNLSPGVYQOGLNAVGTGIGMKAGIOALNDICTHRSSTRSPVNVK 326
Db 215 L-----GNGLGGKGLGELNGPADYEQLGVGVGTGVMKAGIEALNNIGTSDSSTRFTINKD 271
QY 327 DRAMAKEIGQFMDQYPEVFGKPOYKQPGOEKVTDDKSWAKALSCKPDDDDGMPASMEQFN 386
Db 272 DRSIAKEVGQFMDQYPELFGAPQYKTYPSDVKTDTKSWAEALSNDPDDDDGMPASMEQFN 331
QY 387 KAKGMIRPMAGDTGNGNLQARGAG 411
Db 332 KAKGMIRPMAGDTGNGNLQARGAG 354

```

## RESULT 3

```

Q93FY8 PRELIMINARY; PRT: 356 AA.
AC Q93FY8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Harpin.
GN HRPN.
OS Pectobacterium carotovorum subsp. carotovorum.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCC1.
RA Rantakari A.V., Virtaharju O., Vaehnamiko S., Taira S., Palva T.E.,
RT "Type III secretion contributes to the pathogenesis of soft-rotting
RT plant pathogen Erwinia carotovora: Partial characterization of the hrp
RT gene cluster.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF302656; AAK97283.1; -.
SQ SEQUENCE 356 AA; 35534 MW; 154C9D885810DCB8 CRC64;

Query Match 37.4%; Score 840.5; DB 2; Length 356;
Best Local Similarity 48.0%; Pred. No. 4.6e-41;
Matches 198; Conservative 36; Mismatches 114; Indels 59; Gaps 11;

QY 37 LNTSGLCASTMQISIGGAGNGNLLGTSRQNALGNGSALGGLGGNNDTVNOLAGLLT 95
Db 2 LNSLGGAS-LQITI-KAGNGGLFPSSQSGSPSQSAF---GGQRNIAEQLSDIMT 56
QY 96 GMMNMMMSMMGGGLGGLGGLG-----GGLNGLGGSGGLGELSNALNDMLGGSLNLTGS 149
Db 57 TMFMFMMSMMGGG--MSGGLGGLGSSLGGGLGGLG--GGLGGLGSSLGGSLGSALG--GG 111
QY 150 KGCNNTTSTTNSPLDQALGINSNDSTSGTSTSDSSDPMQQLKMFSEIMQSLFGD 209
Db 112 LGG-----ALGAGMAMNPMSMMGS-----LLFSALEDLGG 143
QY 210 GQDGTGSSSGKQPTGEQNAKKYKGVTDALSLGMNGLSQLLNGSLGGGCGGAGTGL 269
Db 144 GMSQOQGLGFLPKPSPEISAVTQGVNDALSAILNGLSQTKGQ-----T 189
QY 270 DGSLSLGGKGLQNLSPGVYQOGLNAVGTGIGMKAGIOALNDICTHRSSTRSPVNVKGDRA 329
Db 190 SPLQNLNGLQGLSGAGAFNLQGLSTLGMVYQKAGLQELANNISTHNDSPTRYFVDEKDRG 249
QY 330 MAKEIGQFMDQYPEVFGKPOYKQPGOEKVTDDKSWAKALSCKPDDDDGMPASMEQFNKAK 389
Db 250 MAKEIGQFMDQYPEVFGKPEYKQDNMTAKODDKSWAKALSCKPDDDDGMPASMEQFNKAV 309
QY 390 GMTRKPMAGDTGNGNLQARGAGSSGIGIDAMACDAINNMAKGLGAA 436
Db 310 GMINTAVAGDTGNTLSARGNGCASLIDAMTGDRIVNMGLKLS 356

RESULT 4
Q9EXP0 PRELIMINARY; PRT: 344 AA.
ID Q9EXP0;
AC Q9EXP0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Harpin protein.
GN HRPN.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;

```



```

QY 291 LGNAVCTGCMKACIOALNDICTHRHSSTRSFVNKGDRAKAEIQPMDOYDEVFGKPOY 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1214 -GNA---GIG-----GDSGSGNGGQ-GDSGSGLGGQPGF 1243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 351 QKGPQGVKRTDDKSWAKLSKPDGCMTPASMEQFNKAKGMIRPMAGDTCNGNLQARGA 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1244 AGPGGG-----KGGAGGNAGTGGTNGSGA 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 411 GGSGLGIDAMMAGDAINNMAKGLG 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1268 GGAGGGGAGGAGISFNSGNGGTG 1292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
*O07224
ID O07224 PRELIMINARY; PRT; 591 AA.
AC O07224;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein Rv0297.
GN Rv0297 OR MTCy63.02.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV.
RX MEDLINE=980295097; PubMed=9634220;
RA Cole S.T., Bresch K., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL: Z96800; CAB09596.1; -.
DR TubercuList; Rv0297; FE_region.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 591 AA; 49056 MW; 4E4675EDD4297F5A CRC64;

```

Query Match 11.3%; Score 255; DB 16; Length 591;  
 Best Local Similarity 26.8%; Pred. NO. 3.8e-07;  
 Matches 123; Conservative 24; Mismatches 168; Indels 144; Gaps 20;

```

QY 41 GLGASTMQISIGGAGNGNLLGTSRQNALGGNSALGLGGGNQNDTVNQLAGLLTHMM 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 GNGGSGAPQACAGACAGAGFFG---NGGNGDGGAGANG-----AGGTAGWFFG 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 MSMMGGGLMG-----GGLGGGLGN-----GLGSGGLGEGLSNALNDML----- 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 FCGNGAGAGIGVAGINGLGGAGGGGNGAGFFGNGGNGMGAGAGVNAVNFGLATPVT 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 141 -----CGSLNTLCSKG--GNNTTSTNSPLDQALGINSTSQNDSTSG-----TDS 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 PAANGNGNLNVGPCTAGGADGANGCAIGAGGAGCGGNGNSTSGGICIAOTGGAGGA 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 169 SDPMQQLKMFSEIMQSLFQDQGTGQSSGKQPTTEGEQNAKKYKGYTDALSLGMLNGL 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 G-----GAGGDGAPGNGNGCGSVEHT-----GATGS-SASGNGA 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 SOLLGNGGLG-----GGCGGNACTGLDSSLCCKKGLONLSGPV 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 TG--GNGGVGAPGGAGNGGHVSGGSVNTAGAGKGGNGGTGGAGGF--GGHGGSVLSGPV 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 287 DYQQLCNAVCTGCMKACIOALNDICTHRHSSTRS----FVNKGDRAKAEIQPMDOY 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 400 GDSNGAGAGDGD---GAGVSATDIAGTGGRGNGHGHGLWTGNGDGGAGGVG----- 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 EVFGKPOVQKGPQGVKRTDDKSWAKLSKPDGCMTPASMEQFNKAKGMIRPM-----AG 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 -----GVG-----GAGAAGATGGHGGD-----GSVNTPIGGSEAG 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 399 DTGNCNIFQARAGSSIGIDAMMAGDAINNMAKGLGAA 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 DGGKGLGGDGGGRGIFG--QFGAGGAGGAGGVGGAGGA 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
Q8VKN3
ID Q8VKN3 PRELIMINARY; PRT; 622 AA.
AC Q8VKN3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN MT0311.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolova M.D., Salzberg S.L.,
RA Dalcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RE EMBL: AS006938; AAK44534.1; -.
RT TIGR: MT0311; -.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR00308; ANTIPEPEZEI.
DR SEQUENCE 622 AA; 52747 MW; B27AD977A4DE867 CRC64;

Query Match 11.3%; Score 254; DB 16; Length 622;  

  Best Local Similarity 27.5%; Pred. NO. 4.6e-07;  

  Matches 125; Conservative 22; Mismatches 172; Indels 136; Gaps 20;

```

```

QY 41 GLGASTMQISIGGAGNGNLLGTSRQNALGGNSALGLGGGNQNDTVNQLAGLLTHMM 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 GNGGSGAPQACAGAGCATGFFG---NGGNGDGGACANG-----AGGTAGWFFG 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 MSMMGGGLMG-----GGLGGGLGN-----GLGSGGLGEGLSNALNDML----- 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 FCGNGAGAGIGVAGINGLGGAGGGGNGAGFFGNGGNGMGAGAGVNAVNFGLATPVT 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 141 -----CGSLNTLCSKG--GNNTTSTNSPLDQALGINSTSQNDSTSG-----TDS 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 PAANGNGNLNVGPCTAGGADGANGCAIGAGGAGCGGNGNSTSGGICIAOTGGAGGA 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 189 SDPMQQLKMFSEIMQSLFQDQGTGQSSGKQPTTEGEQNAKKYKGYTDALSLGMLNGL 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 G-----GAGGDGAPGNGNGCGSVEHT-----CATGS-SASGNGA 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 SOLLGNGGLG-----GGCGGNACTGLDSSLCCKKGLONLSGPV 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 TG--GNGGVGAPGGAGNGGHVSGGSVNTAGAGKGGNGGTGGAGGF--GGHGGSVLSGPV 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 287 DYQQLCNAVCTGCMKACIOALNDICTHRHSSTRSFVNKGDRAKAEIQPMDOYPEVFG 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 GDSNGGAGDGD---GAGVSATDIAGTGGRGNGHGHGL--WINGNGD----- 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 347 KPQYQKGPQGVKRTDDKSWAKLSKPDGCMTPASMEQFNKAKGMIRPM-----AGDTGN 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Db 474 -----GGVGGVGGGAGAGAGHGGDG-----GSVNTPIGGSEAGDGK 515
Qy 403 GNQARGAGSSSLGIDAMMAGDAINNMAKLGIAA 437
Db 516 GGLGGGGGRGIFG-OFGAGGAGGAGGAGGAGGA 548

RESULT 9
Q8VJ19
ID Q8VJ19 PRELIMINARY; PRT; 628 AA.
AC Q8VJ19;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN MT3476.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007154; AAK47814.1; -.
DR TIGR; MT3476; -.
DR InterPro; IPR00104; Antifreeze_1.
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR00308; ANTIFREEZEI.
DR ProDom; PD001223; PE_region; 1.
DR PROSITE; PS00583; PFKB_KINASES; 1; UNKNOWN_1.
SQ SEQUENCE 628 AA; 53828 MW; 9F62B1E7CC2C567E CRC64;

Query Match 11.28; Score 251.5; DB 16; Length 628;
Best Local Similarity 30.68; Pred. No. 6.5e-07;
Matches 118; Conservative 16; Mismatches 155; Indels 97; Gaps 21;

Qy 52 GGAGG--NGLLGTSRQNALGGNSALGLGGNQNDVTNQLAGLLTGMMMMSMMGGGGL 109
Db 320 GGAGGDANGNPANTSIANAGAGGCAAG-GDGGANGGAGGAGGQAASGVGGDGGNGG 378
Qy 110 MGG----GLGGGLNGLGGSGGLGGL-SNALNDMLGSLNTLGSKGNNTTSTNSPLD 164
Db 379 AGGTGTNGHAGGAG-GAGGAGGGRGLVNGGNGNGCN----GAAGNGAIGGTGG--- 430
Qy 165 QALGINTSQNDSTSTGTDSTSDSDPQQLLKMFSEIMQSLFGDQDGTQSSSGGKQP 224
Db 431 -AGGV-PANQGGNSALGT-----QPVGGDGGDGGNGTGG----- 463
Qy 225 TEGQNAYKKVTDALSG-LMGNGLSOLLGNGGLG--GGQGNACTGLDGSLSGKGLQN 281
Db 464 TGGRGDGGSGGAGAGSLWMNGNG--GNGGTGGSGVGVNGGIGDGGAG-GGNATST 520
Qy 282 LSGPVDYQOLGNVGTGIGMKAGIQAALNDIGTHRHSSRFSFNKGDRAKEMIGQFMQY 341
Db 521 SSIPFDAHG-GNG--GAGGDAG-----HGGTGGDGGDGGHAGTGGRGGLL--- 562
Qy 342 PEVFGKPOYKQPGQEVKTTDDKSNKALSKPDDDGMTPASMEQFNKAKGMKRPWAGDTG 401
Db 563 -----AGQHANSNGGGGGTGGAGGTHG-TPGSGN-----AGGTG 596
```

```
Qy 402 NGNLQARGAGSSSLGIDAMMAGDAIN 427
Db 597 TGNADSTNGPGSDG-----LGGDAFN 618

RESULT 10
Q8VK71
ID Q8VK71 PRELIMINARY; PRT; 635 AA.
AC Q8VK71;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN MT1123.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006992; AAK45381.1; -.
DR TIGR; MT1123; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
SQ SEQUENCE 635 AA; 51726 MW; A461A61100C7C5D4 CRC64;

Query Match 11.18; Score 250; DB 16; Length 635;
Best Local Similarity 27.28; Pred. No. 8e-07;
Matches 120; Conservative 23; Mismatches 152; Indels 146; Gaps 20;

Qy 41 GLGASTMQISIGGAGNNGLLGT---SRQNALGGNSALGLGGNQNDVTNQLAGLLTGM 97
Db 293 GSGGHALLMGAGGAGGNGSGGTGGAGGTAGAGGNGGAGGGGT-----GGLLFG- 343
Qy 98 MMSMMMGGLMG-----GGL-----GGGLG--NGLGGSGGLGGLSNALNDMLG 141
Db 344 -----NGAGGOGATAGAGGAGANGVSTNGGTGGNGGIGGTGGSGGAGGNA--GLLG 395
Qy 142 -----GSLNTLGSKGNNTTSTNSPLDQALGINSTQNDDSTSGTSDSDPQQ 194
Db 396 VGGAGGHGASGGADRGAGGTGFISS--DGGAGDGGDGGNGGAGGTGGL----- 444
Qy 195 LKMFSEIMQSLFGDQDGTQSSSGGKQPTEGEQNAKKGVTDALSLGMNG----- 247
Db 445 -----LPGAGNGGPG-GSGGAADIGNGGAGNGGTDGNGCNGSGGGAGSGG 492
Qy 248 -----LSOLLNCGSL--GGCGGN-AGTGLDGSLSGKGLQNLSCPVDYQOLGNV 296
Db 493 DGGAGGNGNAWLFNGGAGGGGKGGNGAGGLGGGSGFLPGL-NGSG-----GDGD 544
Qy 297 TGIKMGAGIQAALNDIGTHRHSSRFSFNKGDRAKEMIGQFMQYPEVFGKPOYKQPGQ 356
Db 545 GGNAPGGVLYNG-----GAGGQSSGGIG-----GPG- 573
Qy 357 EVKTDKSNKALSKPDDDGMTPASMEQFNKAKGMKRPWAGDTGNGNLQARGAGSSSLG 416
Db 574 -----ATGAGGKGGDGG-----DAQLIGDGGNGGNGGAGGTGGTGP 610
Qy 417 IDAMMAGDAINNMAKLGIAA 437
Db 611 PGPGGGGGLGGLLFGQTGTA 631
```

```

RESULT 11
Q8VJ15 ID 08VJ15 PRELIMINARY; PRT; 775 AA.
AC 08VJ15;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN MT3495.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007155; AAK47832.1; -.
DR TIGR; MT3495; -.
DR InterPro; IPR000084; PE_region.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
DR ProSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_3.
SQ SEQUENCE 775 AA; 65448 MW; 5256A18546BDB53 CRC64;

Query Match 11.1%; Score 250; DB 16; Length 775;
Best Local Similarity 28.7%; Pred. No. 9.9e-07;
Matches 96; Conservative 23; Mismatches 101; Indels 114; Gaps 15;

Qy 36 SLNTSGLGASTMQISIGGAGGNNLLGTSRQ-----NAGLGN-SALGLG 79
Dy 446 SAGTGVGAS-----GGTGGNAGLIGAGGCGAGGAGGAGGAGGAGGAGGAG 499
Qy 80 G-----GNQNDTVNQLAGLLTGMMMMMSMGGGLMGGGLG-GGLG-----NLGGSGG 127
Dy 500 GQLYGGGDDGNCAG-----GANIAGNGSDGAGAGGAGGAGGAGGAGGAGGAGG 552
Qy 128 LGEGLSNALNDMLGSLNLTGSKGN-----NTTSTNSPLD 164
Dy 553 AGGNFAGRRADAAGT-----GGDGGNGGNGLLSGNAGAGGAGGAGGAGGAGGAGG 608
Qy 165 QALGINTSONDSTSGTSDSTSDSDPMQQLKMFSETHQSLFGDQDGTQSSSGGKQF 224
Dy 609 GATGGNGGAGGAGTAGTGTSG-----GICGNGGAGCTCGNAG-----646
Qy 225 TEGEQNAYKGVTDAL-----SGLMNGLSQLLNGGLG--GGGGNAGTGLDGSLSGK 277
Dy 647 -----VALSVGSTGLGNGSGGLCGGGGSLFGNGAGVGATGNGGSGTGPAVYGN 701
Qy 278 G-----LQNLSPVDYQOLGNVAGTGLGMKAG 304
Dy 702 GKGGVGAAGLAG-----QINGGGSGSGGAG 730

RESULT 12
Q8VIY9 ID 08VIY9 PRELIMINARY; PRT; 1217 AA.
AC 08VIY9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN MT3615.3.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae.

```

```

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007164; AAK47974.1; -.
DR TIGR; MT3615; -.
DR InterPro; IPR000084; PE_region.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
DR ProSITE; PS00583; PFKB_KINASES_1; UNKNOWN_3.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_3.
SQ SEQUENCE 1217 AA; 94463 MW; 2F5BDC5A30A056E CRC64;

Query Match 11.1%; Score 249.5; DB 16; Length 1217;
Best Local Similarity 25.5%; Pred. No. 1.7e-06;
Matches 112; Conservative 27; Mismatches 148; Indels 153; Gaps 17;

Qy 38 NTSGLGASTMQISIGGA-----GCNNGLLGTSRQAGLGNLSALG-----LGG 80
Dy 731 NPTGIGTGGDGTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 789
Qy 81 GNQNDTVNQLAGLLTGMMMMMSMGGGLMGGGLG-----GGLG--NGLGGSGGLGEG 131
Dy 790 GADADQPGATGG--TGFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 847
Qy 132 LSNALNDMLGGS--LNTLGSKGNNTTSTNSPLDQALGINTSONDSTSGTSDSTSDSS 189
Dy 848 FNSNGGTTGGTGGVGGTGGDGN-----AGTGAGDPKGGTGGTGGTGGTGGSG 894
Qy 190 DPMQQLKMFSETHQSLFGDQDGTGSS-----SGGKQPTGEGQNAVKKGVTDALSG 242
Dy 895 -----GAGSGGANFNGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 934
Qy 243 LMGNLSQLLNGGLG-----GGGAGTGLDGSLSLGGKGLQNLSPVDYQOLGNV 295
Dy 935 TGTGTG-----GTGKGGTGCGAGDSDSAGTGTGCGAGGAGGAGGAGGAGGAGGAG 983
Qy 296 GTGIGWKAGLQALNDIGTHRHSTRSFVNKGDRAKAKEIQFMDQYPEVEFKPOYKGP 355
Dy 984 --GIG-----GDGGGGNGGQ--GDSSGSLGGQPFAGGPG 1015
Qy 356 QEVKTTDDKSNALKSKPDDDGMTTPASMEQFNKAKMTKRPAGDTGNGNLQARGAGSS 415
Dy 1016 G-----KGAGGNAGTGGTNGSGAGGAGG 1039
Qy 416 GIDAMMAGDAINNMAICKLG 435
Dy 1040 QGAGGAGISFSGNSNGTGC 1059

RESULT 13
Q50415 ID Q50415 PRELIMINARY; PRT; 731 AA.
AC Q50415;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PGRS-family protein.
GN RV3388 OR MTV004.46.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]

```





RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AE007163; AKK4790.1; -;
TR	TIGR; MT3612; -;
DR	InterPro; IPR001343; Hemlysn_Ca_bind.
DR	InterPro; IPR000084; PE_region.
DR	InterPro; IPR002173; PfkB.
DR	Pfam; PF00934; PE; 1.
DR	PRINTS; PR00313; CABDNNGRPT
DR	PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
SQ	SEQUENCE 1384 AA; 110839 MW; 124AC8AAADD55EBD CRC64;
	Query Match 11.0%; Score 248; DB 16; Length 1384;
	Best Local Similarity 27.6%; Pred. No. 2.5e-06;
	Matches 112; Conservative 25; Mismatches 161; Indels 108; Gaps
QY	38 NTSGLGASTWQISIGGAGNNGLLGTSR----NQALGGNSALGLGGGNQNDTVNLQGL 93
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
	479 NPQGKGDG--GIGGAGGAGGAATGCGHGAGTGDGCGTGCGNGCTGGVNGADNT 535
QY	94 LTGWMMMSMWGGGGLMG-GLGGGLNGLGSGGLEGLSNALMDLGGSLNTLSKGG 152
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
	536 LN-----PDTFGACEPGCAGGAGGAAGPGCGTCGTCGNGCGNGGNGGNGCN-GGNG 589
QY	153 NTTSTNSPLDALGTNSTSQNDSTSGTDSTSDSDPMQOLLKFSELMQLSFLGDGD 212
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
	590 NAGNNSTNP-----VGEGGAGDGGAGGAGGANG------TAGSQGTGGVGD 635
QY	213 GTQSSSGGKOPTREEONAYKKVTDALSLMGNLSQLLNGCLGGGGGNGAGTLDGS 272
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
	636 GGAGNGGGK--AGTGNSGNFV-DGEAGFSGA-----GGNGGVGGAAGANGTGGSGG 688
QY	273 SLGGKGLQNLSGPVDYQOLNAV-GTGIGMKACTQALNDLTGRHSSTRSFVNKDRA 331
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
	689 NGDGGAGGTCCAG-----GNGIDGTETPDAGTCA-----KG----- 721
QY	332 KEIGQFMDOYPEVFKGPOYKGPGQEVKTDDKSWAKALSPPDDGMTPASMEQFNKAGM 391
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
	722 -----GDG-----GDBGAGGAG-----GNAGGA 739
QY	392 IKRPMAGGTGNGNIQARAGSGSLGIDMAMAGDAINMALGKLGA 437
Db	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
	740 GNGGAGGAGGQ--GNAGCGGAGGA--GGNAVTPGDSGVGKAPHGDAGGS 78

Search completed: January 22, 2003, 16:27:39  
Job time : 39 secs

Result No.	Query	Score	Query			ID	Description
			Match	Length	DB		
1	2050	91.2	403	1	HRPN_ERWAM	Q01099 erwinia amy	
2	828.5	36.9	356	1	HRPN_ERWCA	Q47279 erwinia car	
3	718.5	32.0	340	1	HRPN_ERWCH	Q47278 erwinia chr	
4	248	11.0	463	1	YA68_MYCTU	O53416 mycobacteri	
5	242	10.8	778	1	YQ34_MYCTU	P71933 mycobacteri	
6	241.5	10.7	957	1	Y278_MYCTU	P56877 mycobacteri	
7	238.5	10.6	515	1	Y140_MYCTU	O50594 mycobacteri	
8	238	10.6	1901	1	Y208_MYCTU	O53553 mycobacteri	
9	235.5	10.5	214	1	WA22_MYCTU	O06794 mycobacteri	
10	234.5	10.4	543	1	Y9P1_MYCTU	O50630 mycobacteri	
11	234.5	10.4	603	1	YD35_MYCTU	Q10637 mycobacteri	
12	233.5	10.4	338	1	GRF_ARATH	P27483 arabidopsis	
13	232	10.3	801	1	Y747_MYCTU	O53810 mycobacteri	
14	229	10.2	491	1	YK98_MYCTU	Q10707 mycobacteri	
15	217	9.7	384	1	GRP1_PETHY	P05789 petunia hyb	
16	213.5	9.5	481	1	LORI_MOUSE	MUS MUSCULU	
17	213	9.5	1156	1	GLH4_CAEEL	O76743 caenorhabdi	
18	211.5	9.4	747	1	SPD1_NEPCL	P19837 nephila cla	
19	202.5	9.0	498	1	Y118_MYCTU	O50615 mycobacteri	
20	202	9.0	937	1	HVR1_CANAL	P46591 candida alb	
21	197.5	8.8	344	1	POPA_RALSO	O9rbs0 ralstonia s	
22	196.5	8.7	465	1	GRP2_PHAUV	P10496 phaseolus v	
23	193	8.6	183	1	GRP2_ORYSA	P29834 oryza sativ	
24	191	8.5	252	1	GRP1_PHAUV	P10495 phaseolus v	
25	191	8.5	316	1	LORI_HUMAN	P23490 homo sapien	
26	190	8.5	641	1	EBN1_EBV	P03211 Epstein-bar	
27	187.5	8.3	1387	1	TROP_HUMAN	Q12816 homo sapien	
28	187	8.3	959	1	N100_YEAST	Q02629 saccharomyc	
29	186.5	8.3	1113	1	N116_YEAST	Q02630 saccharomyc	
30	184.5	8.2	419	1	CSP_PLACM	P08676 plasmodium	
31	177.5	7.9	1224	1	PER_DROME	P07663 drosophila	
32	175.5	7.8	734	1	YKR2_CAEEL	P34308 caenorhabdi	
33	174.5	7.8	672	1	PHX5_MOUSE	P08399 mus musculu	

Db 61 TGNMWMHMMGGGLMGGGLGNGLGNGLGSGGLGEGLSNALNDMLGGSINTLGSKGNN 120  
 QY 155 TTTSTNSPDLQALGINSTQNDSTSGTSTSDSPMQQLKMFSEIMQSLFGDQDGT 214  
 Db 121 TTTSTNSPDLQALGINSTQNDSTSGTSTSDSPMQQLKMFSEIMQSLFGDQDGT 180  
 QY 215 QGSSSGGKQPTGEQNAKKGYTDALSGLMNGLSQLLNGGGLGGGGNAGTGLDSSSL 274  
 Db 181 QGSSSGGKQPTGEQNAKKGYTDALSGLMNGLSQLLNGGGLGGGGNAGTGLDSSSL 240  
 QY 275 GKGGLNLSPVDYQOLGNVGTGIGWKAGIOALNDIGTHRHSSSTRSFYKNGDRAMAKEI 334  
 Db 241 GKGGLNLSPVDYQOLGNVGTGIGWKAGIOALNDIGTHRHSSSTRSFYKNGDRAMAKEI 300  
 QY 335 GOFMDQYPEVFGKPOYKQPGQEVKTDDKSWAKALSKPDDGTMTPASMEOFNKAQMIKR 394  
 Db 301 GOFMDQYPEVFGKPOYKQPGQEVKTDDKSWAKALSKPDDGTMTPASMEOFNKAQMIKS 360  
 QY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437  
 Db 361 AMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403

## RESULT 2

HRPN\_ERWCA  
 ID HRPN\_ERWCA STANDARD; PRT; 356 AA.  
 AC Q47279;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Harpin (Harpin-ECC) (Fragment).  
 GN HRPN.  
 OS Erwinia carotovora.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pectobacterium.  
 OX NCBI\_TaxID=554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=71;  
 RX MEDLINE=96405946; PubMed=8810071;  
 RA Cui Y., Medi L., Mukherjee A., Dumenyo C.K., Chatterjee A.K.;  
 RT "The K8Ma mutants of Erwinia carotovora subsp. carotovora strain  
 Ec71 overexpress hrpNec and elicit a hypersensitive reaction-like  
 response in tobacco leaves."  
 RL MOL. Plant Microbe Interact. 9:565-573(1996).  
 CC -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT  
 CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO  
 CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.  
 CC -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; L78934; AAB49733.1; -.  
 KW Hypersensitive response.  
 FT DOMAIN 1 224 GLY-RICH.  
 FT NON\_TER 356 356  
 SQ SEQUENCE 356 AA; 35621 MW; 108B46B9D27F9DE4 CRC64;

Query Match 36.9%; Score 828.5; DB 1; Length 356;  
 Best Local Similarity 48.2%; Pred. NO. 1.5e-43;  
 Matches 196; Conservative 37; Mismatches 115; Indels 59; Gaps 11;

QY 37 LNTSLGLASTMISTGAGGNNGLL-GTSRQNRAGLGGNSALGGLGGNQNDTYNQLAGLLT 95  
 Db 2 LNSLGGGAS-LOITI-KAGNGGLPPSSQNGSPSQSAF---GQGRNIAPQLSDIHT 56  
 QY 96 GMMMMWMHMMGGGLMGGGLG-----GGLGNGLGSGGLGEGLSNALNDMLGGSINTLGS 149

Db 57 TMMPMGMMGGG---MSGCLGCLGSLGGLGGLLGG-GLGGLGSSLSGSGLSALG--GG 111  
 QY 150 KGGNNTTSTNSPDLQALGINSTQNDSTSGTSTSDSPMQQLKMFSEIMQSLFGD 209  
 Db 112 LGG-----ALGAGMNAKMFSAKMG-----LLFSALEDLLGG 143  
 QY 210 GQDGTGSSSGGKQPTGEQNAKKGYTDALSGLMNGLSQLLNGGGLGGGGNAGTGL 269  
 Db 144 GMSQOQGLGFGKQFSSPEISAYTQGVNDALSAILGNGLSOTKGO-----T 189  
 QY 270 DGSLSGKGLNLSPVDYQOLGNVGTGIGWKAGIOALNDIGTHRHSSSTRSFYKNGDRA 329  
 Db 190 SPLOLGNNGLOGSGAGAFNQLGSLGMSVQKAGLOELNINSHNDSPTRFYVDKEDRG 249  
 QY 330 MAKEIGFMDQYPEVFGKPOYKQPGQEVKTDDKSWAKALSKPDDGTMTPASMEOFNKA 389  
 Db 250 MAKEIGFMDQYPEVFGKAEYQKDNWQTAQEDKSWAKALSKPDDGTMKGMKPKKAV 309  
 QY 390 GMTKRPMDGTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 436  
 Db 310 GMIKSAIRGDTGNTLSARGNGGASLGDRAAMIGDRIVNMGELKKLSS 356

## RESULT 3

HRPN\_ERWCH  
 ID HRPN\_ERWCH STANDARD; PRT; 340 AA.  
 AC Q47278;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Harpin (Harpin-ECH).  
 GN HRPN.  
 OS Erwinia chrysanthemi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pectobacterium.  
 OX NCBI\_TaxID=556;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EC16 / AC1450;  
 RX MEDLINE=961172740; PubMed=8589405;  
 RA Bauer D.W., Wei Z.M., Beer S.V., Collmer A.;  
 RT "Erwinia chrysanthemi harpinEch: an elicitor of the hypersensitive  
 response that contributes to soft-rot pathogenesis."  
 RL MOL. Plant Microbe Interact. 8:484-491(1995).  
 CC -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT  
 CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO  
 CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.  
 CC -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; L39897; AAC31978.1; -.  
 KW Hypersensitive response.  
 FT DOMAIN 1 203 GLY-RICH.  
 SQ SEQUENCE 340 AA; 34274 MW; E4D82C23731EF4C5 CRC64;

Query Match 32.0%; Score 718.5; DB 1; Length 340;  
 Best Local Similarity 42.0%; Pred. NO. 6.4e-37;  
 Matches 172; Conservative 41; Mismatches 114; Indels 77; Gaps 10;

QY 47 MQLSI-GGAGNNGLLGTSRQNRAGLGG-NSA---LGLGGGNQNDTYNQLAGLLTGM 101  
 Db 1 MQUTIRAHIGDGLGVSGTGLCAQGLGSLNSAASLGSVDKLSLTDKLTSLTSM--- 57  
 QY 102 SMWGGGGLMGGGLGGLGNGLGSGGLGEGLSNALNDMLGGSINTLGSKGNNTTSTNS 161

Db 58 -----FGALAGLGLAS-SKGLGMSNQLCQSTGN-----GAQASNLLSVPK- 98

QY 162 PLDQALGINSTSDNDSTSGTDTSDSDPMQOLLMKFSEIMQSLFG-----DQDQG 213

Db 99 -----SGDALS-----KMFKALDLDLGHDTVTKLTQSNQ 130

QY 214 TQSSSGGKQPTGEQONAYKKVTDALSLMGNGLSOLLNGGLGGQGGNAGTGLDGGSS 273

Db 131 LANSMLNASQMTQGNNAFSGGVNALSILNGLGQSM-----SGFSQPS 176

QY 274 LGGKGLQNLSPVDYQOLGNNAVCTGIGMKAGIOALNDIGCTHRHSSPRSFVNKGDRAMAKE 333

Db 177 LGAGLGLSGAGANFNQNLGNALGMVGQNALSLNSVSTHVDGNRHFVDKEDRMAKE 236

QY 334 IGQFMDQYPEVFGKPYQKQPGQYKTKDDKSWAKALSKPDDDDMTPASMEQFNKAKGMIK 393

Db 237 IGQFMDQYPEIFGKPYKQDGWSSFKTDDKSWAKALSKPDDDDMTGASMDKFRQAMGMIK 296

QY 394 RPMAGDTGNLQARGAGSSGIGIDAMMAGDAINNWLKGLGAA 437

Db 297 SAVAGDTGNTNLNRGAGGASLGIDAAYVVDKIANNSLGLKLANA 340

RESULT 4

YA68\_MYCTU STANDARD; PRT; 463 AA.

AC O53416;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical PE-PGRS family protein Rv1068c.

GN Rv1068C OR MT1097 OR MTV017.21c.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37Rv;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RA "Deciphering the biology of Mycobacterium tuberculosis from the

RL complete genome sequence.";

RL Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains.";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AL021897; CAAL17184.1; -.

DR EMBL; AE006991; AAK45353.1; ALT\_INIT.

DR TIGR; MT1097; -.

DR TubercuList; Rv1068c; -.

DR InterPro; IPR000084; PE\_region.

DR Pfam; PF00934; PE; 1.

DR ProDom; PD001223; PE\_region; 1.

KW Hypothetical protein; Complete proteome.

FT CONFLICT 218 218 T -> S (IN REF. 2).

FT CONFLICT 235 235 G -> GGGAGIGGADCTKGGDAGCAGGAGGWIHHGGV

FT GGGGTGGCGDGVQGPEDTGAAGGAGG (IN REF. 2).

SQ SEQUENCE 463 AA; 39305 MW; CF5696A7E9593952 CRC64;

Query Match 11.0%; Score 248; DB 1; Length 463;

Best Local Similarity 31.8%; Pred. No. 2.4e-08;

Matches 93; Conservative 17; Mismatches 112; Indels 70; Gaps 13;

QY 42 LGASTMQISIGGAGGNLLGTSTRNAGLGCNSALGGLGGGNQNDTVNQLAGLLTGMMM 101

Db 203 IGAPVAGGAGGAGGTAGLFG-----NGGAGCAGGAGGAGGRGD--GGAGNLSGNGDA 256

QY 102 SMMGGGLMGGLGGGLGNGLGGSGGLGEGLSNALNDMLGGLSLNTLGSKGNNTTSTNS 161

Db 257 GTGGGNGAGNGGNGGSAGWLSNGGTGG-----GGTAGA-GGQGGNG-----NS 301

QY 162 PLDQALGINSTSDNDSTSGTDTSDSDPMQOLLMKFSEIMQSLFGDGDGTQGS--- 218

Db 302 GIDPG---NGGQADTGNAGNGHGS-----AKLFGDGGAGGAGMGST 344

QY 219 -----SGGKQPTGEQONAYKKVTDALSLGMLGNLSQL-----GNGGLGGQG--G 263

Db 345 GGTGGCGGFGGTGGNGCNHAGGAGSGGTAGLLGSGSGGTGGDGGNGGLAGSGAKG 404

QY 264 NAGTGLDSSLGKGLQNLSPVDYQOLGNNAVCTGIGMKAG---IQALNDIG 312

Db 405 NGNGGD-----GGKG-----GDAQLIGNGNGNGGKGGTGLMPGINGTG 445

RESULT 5

YQ34\_MYCTU STANDARD; PRT; 778 AA.

ID YQ34\_MYCTU

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE Hypothetical PE-PGRS family protein Rv2634c.

GN Rv2634C OR MT2712 OR MTCV441.04c.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37Rv;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RA "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence.";

RL Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains.";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

```

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PERS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z80225; CAB02341.1; -
CC EMBL: AE007103; AAK47026.1; ALT_INIT.
CC TIGR: MT0212; -
CC TubercuList; RV2634c; -
CC InterPro: IPR000084; PE_region.
CC Pfam: PF00934; PE; 1.
CC ProDom: PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SQ SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;

Query Match 10.8%; Score 242; DB 1; Length 778;
Best Local Similarity 30.1%; Pred. No. 9.8e-08;
Matches 91; Conservative 19; Mismatches 110; Indels 82; Gaps 14;

QY 41 GLGASTMQISIGAG--GNNLLGTQRNAGLGGNSALGGV-----GNQDNTVQLAGLL 94
Db 353 GNGGNANWFCGAGCGGTGLACTGNCVNPSTANTGANGTDSNGNQV-----404
QY 95 TGMMMMSMMGGGLGGGLG--NGLGSGGLGELSLNALNDMLGSLNLTGSKG 152
Db 405 -----GGNG--GPGFAGVGGEAGVGGGGLGESLDG--NDGTGG-----KGG 443
QY 153 NNTTSTNSPLDQALGINTSTQNDSTSTGTDSTSDSDPMQQLKMFSETMQSLFDGQD 212
Db 444 AGGTAGTGGAGGAGGAGGIGETDGSAGV-----ATGGEQGD 481
QY 213 GTQSSSSGKQTEGE-QNAYKKGVTDAL---SGLMNG---LSQLLNGGLGG-----259
Db 482 GATGGVGVGAGGAGGKCGCOCHNFGVGDATCGDGGICGDCGNCALCAAGNGGTGGAGNGG 541
QY 260 -----GQGNAGTGLDSSLGKGLQNLSPV---DYQQLGNVAVGTGIGMKAGIOALND 310
Db 542 RGMMLNGGAGGAGGTGCT-GGGGAAGFAGVGAGGEGELTDGAGTAEGGTGGLGLGG 600
QY 311 IG 312
Db 601 VG 602

RESULT 6
ID Y278_MYCTU STANDARD; PRT; 957 AA.
AC P56877;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PERS family protein RV0278c precursor.
GN RV0278c OR MT0291 OR MT035.06c.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID-1773;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA DAVIES R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
KL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OshKosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy K., Dodson R., Gwinn M.D., Hart D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL021930; CAAL7353.1; -
CC EMBL: AE006936; AAK44511.1; ALT_INIT.
CC TIGR: MT0291; -
CC TubercuList; RV0278c; -
CC InterPro: IPR000084; PE_region.
CC Pfam: PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 957 HYPOTHEITICAL PE-PGRS FAMILY PROTEIN
FT SIGNAL 40 40 M -> I (IN REF. 2).
FT CONFLICT 158 163 MISSING (IN REF. 2).
FT CONFLICT 807 807 R -> G (IN REF. 2).
SQ SEQUENCE 957 AA; 81905 MW; 71EBAD417FBA47C CRC64;

Query Match 10.7%; Score 241.5; DB 1; Length 957;
Best Local Similarity 25.8%; Pred. No. 1.3e-07;
Matches 117; Conservative 42; Mismatches 149; Indels 145; Gaps 21;

QY 36 SLNTSGLGASTMQISIGAGGNNLLGTSTQNAGLGG-----NSALGLGNGQNDTVNQLA 91
Db 362 TLTAGGIGGA-----GGAGGNAGLLFGSGSGGAGGFGADGGGQGGPGGN-----A 407
QY 92 GLLTGMMMMMSMMGGGLMG--GGLGGGLGNGLGGSGGL-----GGLSNAL---N 137
Db 408 G-----TVFGSGGAGNGGVGGQGFAGGIGAGGTFGLNGNGNGNGGSAVTTGN 458
QY 138 DMLGGSLLTLCGSKGNNNTTSTNSPLDQALGINTSTQNDSTSTGTDSTSDSDPMQQLK 197
Db 459 GGIGGTGVLIGN--GGNGSGGIGAGKAGAGVGSGLLGLLDGFFNAPASTSPHITLQNNVLN 517
QY 198 MFSEIMQS-----LFGDQDQGTQGS-----SSGKQKTEGEQNAKKGVTDALSG--LMNG 247
Db 518 VVNEPFTLGRPLIGNANGTPTCTGADGGAGGWLFGNGANGTPTCTGAAGAGGWLFGNG 577
QY 248 LSQLLGNGGLGG-----GQGNAGTGLDSSGLGKGLQNLGSPVDY 288
Db 578 -----GNGGHGATNTAATATGAGGAGGAGGILFTGTGNGGTG--GIATGAGGTGAGGAGGV 630

```



```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AL022022; CAAL17745.1; -.  
DR Tuberculin; RV3508; -.  
DR InterPro; IPR000084; PE_region.  
DR Pfam; PF00934; PE; 1.  
DR ProDom; PD001223; PE_region; 1.  
KW Hypothetical protein; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN  
FT CHAIN RV3508.  
SQ SEQUENCE 1901 AA; 147627 MW; C7B19D35D0146CD CRC64;  
  
Query Match 10.6%; Score 238; DB 1; Length 1901;  
Best Local Similarity 27.6%; Pred. No. 4.4e-07;  
Matches 124; Conservative 29; Mismatches 180; Indels 116; Gaps 23;  
  
QY 43 CASTWMOISI-CGAGCNGNGLIGTSRONAGLGGNSALGLCGGNQNDVTNOLAGLLTGMMMM 101  
DB 1133 GASTTSINAGGAGNGGTGG-----KGGAGGAGTIGVGGSGGTGGDGDAG----- 1179  
QY 102 SMVGGGGLGGGLG--GGLGN-GLGSGGLG--EGLSNALNDMLGGSNLPLGSKGQNNTRS 157  
DB 1180 --SGGGCGFCAACKAGCGGCGVCGDGGECAGLGLGLSGFDGCGO---GGGGAGGSA 1233  
QY 158 TTNPLDQALGINSTSONDSTSGTSDTSDSDPQQOQLKMFSEIMQSLFG--DGDDGTQ 215  
DB 1234 -----GAGGINGAG-----GAGGTGGAGGCGAP-----ATLIGPGDGDGCG 1270  
QY 216 GSSSCCKOPTGEONAYKKVT-----DALSLGLNGLSOLLNGLG--GGGGNAGTGL 269  
DB 1271 GGIGG-----DGGNAGFAGVPGDGGDGNAGFGAGVP---GDGGIGGTGGAGGAGGAGA 1322  
QY 270 DG-----SSLGKGLQNLSPVDYQO-----LQNA 294  
DB 1323 DGDPSIDGGCGAGGCGGCGKGLNSTGLASASDGGNGGAGGAGNGGDDGFIGS 1382  
QY 295 VGT-GIGMKAGIQAL-NDIGTHRSSTRSFYNKGRAMAKEIGQFMQDYFEYFGRPQYK 352  
DB 1383 GCTGCTGDAVGGLANTCTAGNAGIGAGGCGDGGAGDSCALSDQNGCFAGCGGCG 1442  
QY 353 GPCQEVKTDKSWAKALSKPDDDCMTPASMEQFNKAKMKRPM-----GDTGNCNLQAR 408  
DB 1443 GYGNAGAGGAGGAGGTG-----GTGGAGGCGQNTTGVASEGGAGGCGGCGGCGGI 1494  
QY 409 GAGSSSLGIDAMMAGDAINNALKGLAA 437  
DB 1495 GGAGNAGFGAGVPDGG-----GIGGTGGA 1519  
  
RESULT 9  
WA22_MYCTU  
ID WA22_MYCTU STANDARD; PRT; 914 AA.  
AC 006794;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE WA22 antigen precursor.  
GN WA22 OR RV1759C OR MT1807 OR MTCY28.25C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98205987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gars S., Barry C.E. III, Tekdia F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Heltroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
```

```
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RL complete genome sequence.";  
RN Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolotay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS  
CC SUBFAMILY.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 85.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z95890; CAB09322.1; -.  
DR EMBL; AE007040; -; NOT_ANNOTATED_CDS.  
DR TIGR; MT1807; -.  
DR Tuberculin; RV1759C; -.  
DR InterPro; IPR000084; PE_region.  
DR Pfam; PF00934; PE; 1.  
KW Antigen; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 914 WAG22 ANTIGEN.  
SQ SEQUENCE 914 AA; 74354 MW; F6953C3DDBE6AC8 CRC64;  
  
Query Match 10.5%; Score 235.5; DB 1; Length 914;  
Best Local Similarity 25.8%; Pred. No. 2.9e-07;  
Matches 120; Conservative 38; Mismatches 173; Indels 135; Gaps 20;  
  
QY 41 GLGASTMQISIGGAGGNG-----LLGTSRONAGLGGNSAL-----GLGGNQNDVTNQ 89  
DB 186 GAGGSAMLFAGAGGAGGAGGAANTSIVGIGCTGCTGGNAGMLAGAGGAGGFSFTAGC 245  
QY 90 LAGL-LTGMMMMMMGGGGLMG-----GGLGGGLG--NGLGGSGGLGE-----GLSNA 135  
DB 246 AGGAGGAGGLFTTGGVGGAGGQHTGGAGGAGGAGGLFGAGGAGGAGGFGDHTLCTGGA 305  
QY 136 LNDMLGCSNLTLGSKGNNNTSTNPLDQALGINSTSONDSTSDTSDSDSDPQQOQL 195  
DB 306 GGDGGGGGLFGAGGCGGAGGCGGLTTGGA-AGNCGNAGTSLGAGGAGGAGGTGGAGG 359  
QY 196 LKMFSEIMQSLFGDGDGTGGS-----SSGKQPTGEQNAKKYKVTDALSLGLMNG 247  
DB 360 -----TVFGGGKAGCAGCGNAGMLFCGCGCGGTGG-----FGAAGCGGAGGGS 404  
QY 248 LSOLLNGGGLGGGOG-----GNAGTGLDGSLSLGGKGLQLNLSGP 285  
DB 405 AGMLSGSGSGGAGGSGGFPAGTAAGAGGAGGAGGAPGLIGNGNGNGGEGSGCTGCGVAG- 463  
QY 286 VDYQQLGNVAGTGTGMKAGTQA-----LNDIGTHRSSTRSFVNRKGDRA 331  
DB 464 -----GNVGLNGGEGGIGLAGSGFGGFGGLLLGADGYNAPESTSPWHN-----LQ 512  
QY 332 KEIGQFMQDYPEVF-GKPYQKPGQEVKTDKSWAKALSKPDDDCMTPASMEQFNKAK 390  
DB 513 QDILSPINEPTALTGRPLTQNG-----DSCTPTGCTDD-CCAGC 550
```





```
CC      Or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z73902; CAA98089.1; -.
DR      EMBL; AE007010; AAK45630.1; ALT_INIT.
DR      TIGR; MT1367; -.
DR      TubercuList; Rv1325c; -.
DR      InterPro; IPR000084; PE_region.
DR      Pfam; PF00934; PE; 1.
DR      ProDom; PD001223; PE_region; 1.
*KW      Hypothetical protein; Repeat; Signal; Complete proteome.
FT      SIGNAL          1   30
FT      CHAIN           31   603
FT                                     POTENTIAL.
FT      HYPOTHETICAL PE-PCRS FAMILY PROTEIN
FT      Rv1325C.
FT      DOMAIN          114   603
FT      CONFLICT        132   132      G -> V (IN REF. 2).
FT      CONFLICT        135   135      G -> D (IN REF. 2).
FT      CONFLICT        337   337      G -> D (IN REF. 2).
FT      CONFLICT        508   508      L -> F (IN REF. 2).
FT      CONFLICT        508   508      L -> F (IN REF. 2).
SQ      SEQUENCE        603 AA; 49575 MW; 4F9BCB82B07AE964 CRC64;

Query Match          10.4%; Score 234.5; DB 1; Length 603;
Best Local Similarity 25.9%; Pred. No. 2.1e-07;
Matches 116; Conservative 35; Mismatches 142; Indels 155; Gaps 24;

QY      20 LFLIISHSHAQNSRMSLN-----TSLGASTMQISTIGGACGNN--GLLGTSRONAGLGG 72
Db      247 LFEVLAAGGAGGTGLSVNGTGGTGGTGGGLFSSNGGAGGAGGFGVSGSAGGNGGTGG 306
QY      73 NSAL-----GLGGGNQNTVQLAGLLTGMMMMMMGGGG-----LMG-CGLGQ-----G 117
Db      307 DGGTPTNCGCTGCTGCTGNQLVGGEG-----ACGACGNAGTLFCAGGIGTGGTGG 359
QY      118 LG-----NGLGGGGLGGLSALNDMLGSLNLTGSKGNNNTTSTTNSPLDQALGINSTS 173
Db      360 LGAPDPGTGKGGVG-GIGCA-----GALPGCGAGGTG-----GFGASSAD 401
QY      174 QNDSTSTGTDSTSDSPMQQLKMFSEIMQSLFGDGDGTGCGSSSGKQKOPTGEONAYK 233
Db      402 QMAGGIGGGGGGAA-----KLIGDGGAG-----GTGG----- 430
QY      234 KGVTDALSGLNGLSQLLNGGLG-----GGQGNAGTCLDSSLGKGLONLSQPDVY 288
Db      431 ----DSVRGAAGSG-----GTGGTGLLDGAGGAGGTGIEFGSVGGAGGAG----- 474
QY      289 QQLGNAGV-TGIGMKAGIQALNDICTHRHSSTSFVNKGDRAKAIQFMQYPEVFGK 347
Db      475 ---GNAAGLSGAGGAGGAGGEE-----TAGDQAGGNAGLL----- 508
QY      348 POYKPGQGEVKTDDKSWAKALSKPDDGDMTPASMEQFNKAKGMIRKRPWAGDTGNGLQA 407
Db      509 ----NGDG-----GAGGAGGLGIAGDGG-----NGGKGG-KAGMVGNGDGG----- 544
QY      408 RCAGGSSGLCIDAMMAGDAINNMALKLG 435
Db      545 -GAGGASVAVGNGVGGSGGNATLIGNG 571

RESULT 12
GRP_ARATH
ID      GRP_ARATH      STANDARD;          PRT;          338 AA.
AC      P27483.
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      01-AUG-1992 (Rel. 23, Last annotation update)
DE      Glycine-rich cell wall structural protein precursor.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia.

ID      GRP_ARATH      STANDARD;          PRT;          338 AA.
AC      P27483.
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      01-AUG-1992 (Rel. 23, Last annotation update)
DE      Glycine-rich cell wall structural protein precursor.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia.
```

```
RX      MEDLINE=92003708; PubMed=1912511;
RA      Quigley F., Villiot M.L., Mache R.;
RT      "Nucleotide sequence and expression of a novel glycine-rich protein
RL      gene from Arabidopsis thaliana.";
RL      Plant Mol. Biol. 17:949-952(1991).
CC      -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC      -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X58338; CAA41249.1; -.
DR      FIR; S17732; KNU0.
KW      Cell wall; Structural protein; Repeat; Signal.
FT      SIGNAL          1   20      POTENTIAL.
FT      CHAIN           21   338      GLYCINE-RICH CELL WALL STRUCTURAL
FT                                     PROTEIN.
FT      DOMAIN          21   338      GLY-RICH.
FT      SEQUENCE        338 AA; 23891 MW; 046A6E8C1A4E89EB CRC64;
SQ      SEQUENCE        338 AA; 23891 MW; 046A6E8C1A4E89EB CRC64;

Query Match          10.4%; Score 233.5; DB 1; Length 338;
Best Local Similarity 29.9%; Pred. No. 1.3e-07;
Matches 80; Conservative 14; Mismatches 77; Indels 97; Gaps 10;

QY      41 GLGASTMQISTIGGACGNNLLGTSRONA--GLGCNSALGLGGGNNONTVNLGLLGTGMM 98
Db      59 GLG-----GGAGGGGIGGGAGGAGGGGLGGAGGGGLGGH-----GGGIGG-- 100
QY      99 MMSMMGGGGLMGGGLGGGLNGLGG--SGLGEGLSNALNDMLGSLNLTGSKGNNNTT 156
Db      101 -----GAGGAGGGGLGGGGHGGGGGGGGGGGGGGGGTGGGAGGGGGGLGGH-- 152
QY      157 STTNSPLDQALGINSTSQNDSTSGTSDSTSDSPMQQLKMFSEIMQSLFGDGDGTG 216
Db      153 -----GGGIGGAG-- 161
QY      217 SSSGGKOPTGEONAYKGVTDALSLGMLNGLSOLLGNGSLGGGGGOGNAGTGLDSSSLGG 276
Db      162 GGAG-----GLGGHGGGGIGGGAGGSGGGGL-----GGGIGGAGGGAGCG--GGAGCG 209
QY      277 KGLQNLSPGVYDQQLGNNAVCTGICMKAG 304
Db      210 GGLGCGHG-----GCGFGGAGGGGLGGGAG 233

RESULT 13
Y747_MYCTU
ID      Y747_MYCTU      STANDARD;          PRT;          801 AA.
AC      O53810.
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical PE-PCRS family protein Rv0747 precursor.
OS      Mycobacterium tuberculosis.
OC      bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC      Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=H37Rv;
RX      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
```



```
FT CONFLICT 312 312 C -> CC (IN REF. 1).
SQ SEQUENCE 491 AA; 41979 MW; 12C8630C59CA0C13 CRC64;

Query Match 10.2%; Score 229; DB 1; Length 491;
Best Local Similarity 29.2%; Pred. No. 3.7e-07;
Matches 106; Conservative 22; Mismatches 149; Indels 86; Gaps 17;

QY 43 GASTMQISIGAGGNNLLGTSRQAGLGGNSALGLGGGNQNDTVNQLAGLL----- 94
-bb 145 GDSFSPGVAGAGGSAGLIG-----NCGRGCGAPGGAGCN-----GGLGGLLLNGGGAGG 195
QY 95 --TGMMMMMSMMGGGLMGLGG-----GLGNLGGSGGLG 129
Db 196 VGRGDNVGLGAGGGGGGGLGAGLIGHGAGGNGCGCHGCGSKAGCGSGGGGPG 255
QY 130 ECLSNALNDMLGGSILNLCGCGNNNTTNSPLDQALGINSTS--QNDSTSGT--DSTS 186
Db 256 Q--FGGAGGLLYNGGAAGS--GGNGGDAGTGVSSDGFAGLGGSGRGDAGLIGVGGGGG 312
QY 187 DSSDP--MQQLLKMFE-----IMOSLFGDQDDTQGSSSGCKQPTGEGQNAVKKCVTDA 239
Db 313 NCGDPLGLCARLFQVCSRGDGGVGGWLYGDGGGGDGGNGG--LPIGSTNA----- 362
QY 240 LSLGMLGNGLS--OLLGNGLGGGGGNNAGTGLDSSLGKGLQNLGSPVDYQQLGNAYGTG 298
Db 363 ----GNGGSARLIGNGAGGSGSGAPGVSOGGAGNPGNCGGNGGVWYNGC--GAG 416
QY 299 ICMKAGIOALNDI-----CTHRHSSTRSFVNKCDRAKEIGOFMDOYPEVFGKPYOK 352
Db 417 GAAGGGPGMMTTSFPGPGGVGGHGGTALLFGDGGAGGAGAGG-----P-----GTFDGA 468
QY 353 GPG 355
Db 469 GPG 471

RESULT 15
GRP1_PETHY STANDARD; FRG; 384 AA.
AC P09780;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Glycine-rich cell wall structural protein 1 precursor.
GN GRP-1.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RA Condit C.M., Meagher R.B.;
RT "A gene encoding a novel glycine-rich structural protein of petunia.";
RL Nature 323:178-181(1986).
CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -!- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.
CC -!- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF
CC FORMING A BETA-SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.
CC -!- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED
CC FAMILIES OF REPEATS, P1 AND P2, EACH REPEAT CONTAINING ABOUT 40
CC AA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04335; CAA27866.1; -.
```

```
DR PIR; A26099; A26099.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 384 GLYCINE-RICH CELL WALL STRUCTURAL
FT PROTEIN 1.
FT DOMAIN 41 384 GLY-RICH.
SQ SEQUENCE 384 AA; 28777 MW; C8541C549417D18C CRC64;

Query Match 9.7%; Score 217; DB 1; Length 384;
Best Local Similarity 27.5%; Pred. No. 1.5e-06;
Matches 86; Conservative 18; Mismatches 129; Indels 80; Gaps 11;

QY 1 MGFFLESQMPSPFLVLTLLFLIISHSHQAQNRMSLNTGCLCASTMQISIGAGCANNCL 60
Db 9 ICLLLFSSI--FFELTAITL-----ADKLEESRWG--NDNGCGFRRRGCGGGRGGRGPS 60
QY 61 LGTSR--ONAGLGGNSALGLGGGNQNDTVNQLAGLLTGMMMMMSMMGGGLMGLGGGLG 119
Db 61 FGRGRAGGGFGGGAGGAGGG-----LGGGGGLGGGGAGGG 98
QY 120 NGLGSGGLGCEGLSNALNDMLGSLNLTGSKGNNTTSTNSPLDQALGINSTSQNDST 179
Db 99 GGLGGGGAGGGFGGGAGGAGGGGLGGGGGLGGGGGGG-----AGCGGVCGGAGSC 150
QY 180 SGTDSSTSDSPMQQLLKMFEIMQSLFCDDQDCTQCSSSGCKQPTGEGQNAVKKCVTDA 239
Db 151 GG-----FCAGGCGVGGGAGAGG---GVGGGGGFGGGGGG 182
QY 240 LSLGMLGNLSQLLNGGLGGGGGNNAGTGLDSSLGKGLQNLGSPVDYQQLGNAYGTGI 299
Db 183 VGGGSGHG-----GFGAGCGVGGGAGGGL--CGGVGGGGGGGGSGG-----GGGIGGGS 299
QY 300 GMKAGIQALNDIG 312
Db 230 GHGGFGAGGGVG 242

Search completed: January 22, 2003, 16:26:56
Job time : 17 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:07 ; Search time 21 Seconds  
(without alignments)  
2000.512 Million cell updates/sec

Title: 693\_11\_3  
Perfect score: 2248  
Sequence: 1 MGFFLFQMPSPFLVSTLLI.....DAMMAGDAINNMALGKLGA 437

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2050	91.2	403	2 T08471	harpin - Erwinia a
2	264.5	11.8	1489	2 D70807	hypothetical glyci
3	255	11.3	591	2 B70523	hypothetical glyci
4	249	11.1	731	2 C70974	hypothetical glyci
5	248	11.0	463	2 B70893	hypothetical glyci
6	248	11.0	1381	2 E70806	hypothetical glyci
7	247.5	11.0	1079	2 B70807	hypothetical glyci
8	247	11.0	588	2 F70971	hypothetical glyci
9	246	10.9	1538	2 B70846	hypothetical glyci
10	245.5	10.9	923	2 E70820	hypothetical glyci
11	245.5	10.9	1660	2 A70869	hypothetical glyci
12	242	10.8	778	2 F70963	hypothetical glyci
13	241.5	10.7	957	2 D70835	hypothetical glyci
14	238.5	10.6	515	2 B70663	hypothetical glyci
15	238.5	10.6	882	2 B70812	hypothetical glyci
16	238	10.6	1901	2 F70806	hypothetical glyci
17	236.5	10.5	714	2 A70807	hypothetical glyci
18	236.5	10.5	853	2 A70896	hypothetical glyci
19	235.5	10.5	914	2 D70987	hypothetical glyci
20	235	10.5	439	2 D70954	hypothetical glyci
21	234.5	10.4	543	2 F70726	hypothetical glyci
22	234.5	10.4	603	2 A70770	hypothetical glyci
23	234	10.4	2174	2 E95965	hypothetical glyci
24	233.5	10.4	338	1 KNMU	glycine-rich cell
25	233.5	10.4	434	2 E70768	hypothetical glyci
26	233.5	10.4	615	2 H70589	hypothetical glyci
27	232	10.3	801	2 F70824	hypothetical glyci
28	232	10.3	837	2 E70835	hypothetical glyci
29	229.5	10.2	1329	2 E70917	hypothetical glyci

30	228.5	10.2	667	2 A70893	hypothetical glyci
31	228	10.1	783	2 E70824	hypothetical glyci
32	227	10.1	496	2 H70839	hypothetical glyci
33	226.5	10.1	491	2 D70916	hypothetical glyci
34	226.5	10.1	576	2 A70900	hypothetical glyci
35	226	10.1	694	2 F70868	hypothetical glyci
36	225.5	10.0	1011	2 F70620	hypothetical glyci
37	225	10.0	749	2 A70812	hypothetical glyci
38	224	10.0	1306	2 A70934	hypothetical glyci
39	223	9.9	562	2 B70953	hypothetical glyci
40	222	9.9	396	2 T49109	glycine-rich prote
41	222	9.9	767	2 E70895	hypothetical glyci
42	221.5	9.9	639	2 D70931	hypothetical glyci
43	220.5	9.8	484	2 F70846	hypothetical glyci
44	220	9.8	532	2 F70580	hypothetical glyci
45	219.5	9.8	741	2 G70917	hypothetical glyci

ALIGNMENTS

RESULT 1

T08471  
harpin - Erwinia amylovora  
C:Species: Erwinia amylovora  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
C:Accession: T08471  
R:Laby, R.J.; Kim, J.F.; Beer, S.V.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: Z16433  
A:Accession: T08471  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-403 <LAB>  
A:Cross-references: EMBL:M92994; NID:g4309677; PID:g4309678  
A:Experimental source: strain 321  
C:Genetics:  
A:Gene: hrpN

Query Match 91.2%; Score 2050; DB 2; Length 403;  
Best Local Similarity 99.0%; Pred. No. 6.7e-123;  
Matches 399; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	35	MSLNTSGLCASTMQISIGGAGGNNLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL	94
Db	1	MSLNTSGLCASTMQISIGGAGGNNLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL	60
Qy	95	TCMWMMSMMGGGLMGGGLGGGLGNGLGGGGLGGLSNALNDMLGGSNTLTGSKGNN	154
Db	61	TCMWMMSMMGGGLMGGGLGGGLGNGLGGGGLGGLSNALNDMLGGSNTLTGSKGNN	120
Qy	155	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDDPMQQLKMFSEIMQSLFGDGQDGT	214
Db	121	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDDPMQQLKMFSEIMQSLFGDGQDGT	180
Qy	215	QSSSGGKOPTGEQNAKGYTDALSGLMGNGLSQLLNGGLGGGGGAGTGDLGSSSL	274
Db	181	QSSSGGKOPTGEQNAKGYTDALSGLMGNGLSQLLNGGLGGGGGAGTGDLGSSSL	240
Qy	275	GGKGLQNLSPDYDQOLGNVAVGTGIGMKAGIOALNDIGTHRHSSSTRSFVNKGDRAMAKEI	334
Db	241	GGKGLQNLSPDYDQOLGNVAVGTGIGMKAGIOALNDIGTHSSDSTRSFVNKGDRAMAKEI	300
Qy	335	GQFMQDYPEVFGKPYQKPGQGEVKTDDKSWAKALSKPDDGTMTPASMEQFNKAGMIKR	394
Db	301	GQFMQDYPEVFGKPYQKPGQGEVKTDDKSWAKALSKPDDGTMTPASMEQFNKAGMIKS	360
Qy	395	PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437	
Db	361	AMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403	

RESULT 2

D70807  
Hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D70807  
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:ACCESSION: D70807  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
a:molecule type: DNA  
A:Residues: 1-1489 <COL>  
A:Cross-references: GB:AL020223; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g292445  
A:Experimental source: Strain H37RV  
C:Genetics:  
A:Gene: Rv3514  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.98; Score 264.5; DB 2; Length 1489;  
Best Local Similarity 26.18; Pred. No. 5e-09;  
Matches 116; Conservative 27; Mismatches 149; Indels 153; Gaps 16;

QY 37 LNTSGLGASTMQI-----SIGGAGNNGL-----LGTSRON 67  
DB 955 LNTDGLSSATSCCTCCTGCKGCTGCAGDSDSAGTGTGGAGNAGAGGLANTGGTAGN 1014  
QY 68 AGLGGSNALGLGGNNDVTNQL-----AGLLTGMNMMMSMMGGLMGGGLGGGLGNG 121  
DB 1015 AGTGGDGGGQ-GNGGQDSCSLGGLGQPGFAGGAGGKGGAGGSGAGGTNGSGGAGGAG-G 1072  
QY 122 LCGSCGLCECLNALNDMLGGS--LNTLCKGCGNTTSTTNSPLDQALGINTSQNDST 179  
DB 1073 QGGAGGAGISFSMGNSNGTGTGGVGGTGGDCGN-----AGTCACTPKCGCT 1119  
QY 180 SGTDSSTDSDPQQLLKMFEIMQSLFGDGDGTGSS-SGGKOPTGEQNAKKGVTD 238  
DB 1120 GTTCTGCGSG-----CAGSGGANFNCGTGTGTGTGTGGKGGMG 1159  
QY 239 AGLSLGNGLSQLLNGNLG-----GGCGNAGTGLDSSSLGKGLQNLGSPVDYQ 290  
DB 1160 GTAGDGGPGDGG--GNAGVGKGTGTNGSGSGTGTGTGGAGNAGAGGLANTGGTA--- 1213  
QY 291 LGNAVGTGICMKACIOALNDICTHRSFVNKGDRAKAEIKGFMDQYDEVFKPKPY 350  
DB 1214 -GNA---GTG-----GDGCGGNGGQ--GDSGSLGCGPGF 1243  
QY 351 QKGPQGEVKTDDKSWAKSKLPDDGTFASMEQFNKAKMIKREPMAGDTGNGLQARGA 410  
DB 1244 AGPGGG-----KCGAGGNAGTGTGNGSGA 1267  
QY 411 GSSSLGIDAMMAGDAINNMLGKLIG 435  
DB 1268 GAAGAGGAGGAGCAGTSGNSGNGG 1292

RESULT 3  
B70523  
Hypothetical glycine-rich protein Rv0297 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 21-Jan-2000  
C:Accession: B70523  
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: B70523

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A;Residues: 1-591 <COL>  
A:CROSS-references: GB:Z96800; GB:AL123456; PIDN:g3761800; PIDN:CAB09596.1; PID:e312165  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Kv0297  
C:Superfamily: unassigned collagens

Query Match 11.3%; Score 255; DB 2; Length 591;  
Best Local Similarity 26.88; Pred. NO. 6.9e-09;  
Matches 123; Conservative 24; Mismatches 168; Indels 144; Gaps 20;

Oy 41 GLGASTWQISICGAGGNLGLTSTRONAGLCNSALGLGGCNQNDTVNQLAGLITGMMM 100  
| | | | |  
Db 141 NGSGSAPQGAGGAAGPFG---NGGNGDGGAGANG-----AGGTAGVFFG 187  
| | | | |  
Oy 101 MSMCGGCLMC-----CGLGCCLEN-----GLGGSGGLEGSNALNDML----- 140  
| | | | |  
Db 188 FGGNGAGGIAGVINGSLGGAGGDGNGAGFFGNGMGGMGAGAAGVNPNPLATPV 247  
| | | | |  
Oy 141 ----GGSLNTLSKG--GNNTTTSPDLDAALINSTSONDDSTSG----TDSTSDS 188  
| | | | |  
Db 248 PAAMGCNLNVDPFACGCCADGCANSAIQACACAGCDGNASTGGIGIAQFGGAGGA 307  
| | | | |  
Oy 189 SDPMQLKMPSEIMSQLIFGODGTGSSGGKOPTGEQNAYKKGVTDALSGLMGNL 248  
| | | | |  
Db 308 G-----GAGDGAFPGNGNGSVYEHT-----CATCS-SASGNGA 342  
| | | | |  
Oy 249 SOLLONGGLQ-----CGCGNAGTGELDGSSLGCKGLQNLGPV 286  
| : - - - - -  
Db 343 TG--GGGVGAPGAGGNGHVGSGSVNTAGACGKGNGCTGGAGGP-GGHGGSVLSCP 399  
| | | | |  
Oy 287 DYQQLGNAGVTGIMKAGIQALNDIGTHRHSSTRS----FYNKGDRAMAKEIQQFMDOYP 342  
| | | | |  
Db 400 GDSCNGGAGGDG---GAGVSATDTAGTGRGNGHGHLMTGNGCDGAGGVG----- 449  
| | | | |  
Oy 343 EVFKPKYQKPGQGOEVKTDDKSWSAKALS KPDDDDTPASMEOFNKA GKMKRPM----AG 398  
| | | | |  
Db 450 -----GVG-----GAGAAGIGHGGD-----GSVNTPIGGSSEAG 480  
| | | | |

Oy 399 DTGNGNIQARAGAGSSIGIDAMAGDAINNMAKCLKAA 437  
| | | | |  
Db 481 DGGKGGLLGGDGGRGRIFG--OFAGGAGGAGGVGGAGGA 517  
| | | | |

RESULT 4  
C70974  
hypothetical glycine-rich protein Rv3388 - Mycobacterium tuberculosis (strain H37RV)  
C/Spectes: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C/Accession: C70974  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Connor, K.; Davies, K.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: SGARES, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A/Reference number: A70500; UID:98259987; PMID:9534230  
A/Accession: C70974  
A>Status: preliminary;  
A/Molecule type: DNA  
A/Residues: 1-731 <COL>  
A:CROSS-references: GB:AL009198; GB:AL123456; MID:g3242262; PIDN:CAA15773.1; PID:g266  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv3388  
C:Superfamily: elastin

Query Match	11.1%	Score 249;	DB 2;	Length 731;
Best Local Similarity	28.7%	Pred. NO. 2.le-08;		
Matches 96:	Conservative	23;	Mismatches 101;	Indels 114;
				Gaps 15;

A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: B70523

```
QY 36 SLNTSGLCASTMQISIGGAGNNGLLGTSRQ-----NAGLGGN-SALGLG 79
Db 402 SAGTGGVGCAS-----GCTGGMNMLLIGAGHGGAGGAGGNTGGVGGNGAGGAGGAG 455
QY 80 G---GNQNDTVNQLAGLLTGMMNMMWGGGLMGGGLG-CGLG-----NGLGGSGG 127
Db 456 GOLYNGGDDGGAG-----GANIAGNGSGDGAAGHGGAGGSARLIAGHGGGGG 508
QY 128 LGEGLSNALNDMLGSLNLTGSKGN-----GGDGGNGGCLLSNACAGHGGAGGSGSTATTTGTPT 564
Db 509 AGGNTARRADAIAGT---GGDGGNGGCLLSNACAGHGGAGGSGSTATTTGTPT 564
QY 165 QALGINTSQNDSTGTDSTSDSPMQQLKMFSEIMQSLFGDQDGTGSSSSGKQP 224
Db 565 GATGGNGGAGGAGTAGTGG-----GIGNGGAGGTGGNAG-----602
QY 225 TEGSONAYKKGVTDAL-----SGLMGNCLSOLLNGGLG---GGGGNAGTGLDGSLLGK 277
Db 603 -----VALSVGSGTGLGNGSGGLGGGSLFNGGAGGVGATGGNGGSGGIGPASVGGN 657
QY 278 G-----LQNLSPVDYQQLGNVAGTGTGMKAG 304
Db 658 GKGGVGAAGGLAG-----QIGNGSGSGSGGAGG 686

RESULT 5
B70893
hypothetical glycine-rich protein Rv1068c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70893
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70893
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-463 <COL>
A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17184.1; PID:g289670
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1068c
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 11.0%; Score 248; DB 2; Length 463;
Best Local Similarity 31.8%; Pred. No. 1.5e-08;
Matches 93; Conservative 17; Mismatches 112; Indels 70; Gaps 13;

QY 42 LGASTMQISTGAGGNNLLGTSRQNALGGSNALGLGGGNNQNDTVNQLAGLLTGMMMM 101
Db 203 IGAPVAGGAGGAGGTAGLFG-----NGCAGGAGGAGGAGGRGDD---GGSAGWLSCNGGDA 256
QY 102 SMMGGGLMGGGLGGLGCLGGLGGLGEGLSNALNDMLGSLNLTGSKGNNNTTSTNS 161
Db 257 GTGGGGNAGNGGCGSAGWLSNGGTGG-----GGTAGA-GGCGNG-----NS 301
QY 162 PLDQALGINTSQNDSTGTDSTSDSPMQQLKMFSEIMQSLFGDQDGTGSS---218
Db 302 GIDPG---NGGQADTGNAGNGHGSSA-----AKLFGDGGAGGAGGCMST 344
QY 219 -----SGGKOPTGEQONAYKKGVTDALSLGMLGNLSQL-----GNGGLGGGQ---G 263
Db 345 GGTGGGGFGGCTGGNGNGHAGGAGGSGGTAGLLGSGSGGTGDDGNGGGLGAGSGAKG 404
QY 264 NACTGLDSSLLGKGLQNLSPVDYQQLGNVAGTGTGMKAG---IQALNDIG 312
Db 405 NNGNGGD-----GGRG-----GDAQLTGNGGNGGKGTGLMPGINTGT 445
```

```
RESULT 6
E70806
hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70806
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70806
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1381 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17744.1; PID:g292
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3507
C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.0%; Score 248; DB 2; Length 1381;
Best Local Similarity 26.8%; Pred. No. 5.1e-08;
Matches 117; Conservative 24; Mismatches 180; Indels 116; Gaps 16;

QY 41 GLGASTMQISTGAGGNNLLGTSRQNALGGSNALGLGGGNNQNDTVNQLAGLLTGMMMM 100
Db 822 G:CAOQTITVPCFNGNAG-DGNGNGNAGAGNGSGDFGNT-----864
QY 101 MSMMGGGLMGGGLGGLGGLGGLGEGLSNALNDMLGSLNLTGSKGNNNTTSTN 160
Db 865 -SGASGGNGNAGTAGSGAGGTGCTGLSGGNGGNGNGN---GGDGGNAGHTVG 919
QY 161 SPLDQALGINTSQNDSTGTDSTSDSDP---MQQLKMFSEIMQSLFGDQDGTGSS 218
Db 920 AQFVPATSLTPNGAGGNGGTGSGNGGAPGAPGPTTGGNAGSQGIGDGGNGDGGK 979
QY 219 SGK-----QPTGEONAYKKGVTDALSLGMLGNLSOLLNGGLG-----258
Db 980 GGDGADAVNVFMPTPQ-----AATGTAGAGDPTGGNGGPGTGPSPVMAPPP 1029
QY 259 -----GGCGNAGTGL---DGSLGKGLQN-----LSGPVDYQQLGNVAGTGTGM 301
Db 1030 TPTVQVQGGDGGAGGTGTNANDGTATGKGEGGVGSLGP-----GGNGGTG---1080
QY 302 KAGIQALNDIGTHRSSTRSFVNKGDRAKKEICQFMDQYPEVGRPQYKQPGQYKTD 361
Db 1081 -----GNASATGTNCVANAGNGKGGDGGQF-----GAGGNGGAGGSV--T 1119
QY 362 DYSWAKALSKPDDGTPASMEQFNKAKMIKRPMAAGTGNGLQARGAGGSLGDMM 421
Db 1120 DGSAGTAGNGNG-----NATNGTI---AQCPAGNGSAGGKGGDGNIAAGA 1166
QY 422 AGDAIN--NMAIGKLG 436
Db 1167 TGTAGNGGNGNDGA 1183

RESULT 7
B70807
hypothetical glycine-rich protein Rv3512 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
```

A:Accession: B70807  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1079 <COL>  
A:Cross-references: GB:AL022022; GB:ALL23456; NID:g3261554; PIDN:CAAL7749.1; PID:g292444  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3312  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 11.0%; Score 247.5; DB 2; Length 1079;  
Best Local Similarity 26.0%; Pred. NO. 4.2e-08;  
Matches 119; Conservative 26; Mismatches 170; Indels 143; Gaps 16;

Qy 32 NSRMSLNTSGLGASTMQLSTIGGAGNGNLLGTSRONAGLGGNSALGLGGGNDNTVNQLA 91  
Db 629 NRNSGNGTCGACGNGCGGANGAGGAGCGGGGTGNGGAGGADGAGNGGNGTNGNGN 608  
Qy 92 GLLTGMMMMMMGGGGL--MGGLGGLGNCGLGCGGGLGEGLSNALNDMLGGLSNT--- 146  
Db 689 G-----GNGGIAGMGGNGGAGTGSNGGNGGSG--GNGGNAGMGNGSGTGS 733  
Qy 147 LQSKCQNTTSTNSPLDQALGINST-----SQNDSTSTGTDSTSDSDPMQQLL 196  
Db 734 DGGAGGGAAGTGGTGGDG-GLTGTTGGTGGSGGTGGDGGNGGNGADNTAN----- 783  
Qy 137 KMFSEIMQSLFGDQDGTQGSSSG-----GKQTEGEQNAKKG----- 235  
Db 784 ----WTAAGDGGCGDGGFGGACAGCGCLTAGANGTGGGCGAGCGGCNCAICGHP 838  
Qy 236 VTDAUSGLMNGLSQLLNGGGL-----GGGNGAGTGLDGSSLGK 278  
Db 839 LTTDDPGNGGTG-----GNGGTGTGGAGTGLGGLGTTGGDGGNGGNGTGGGEGVGGAG 893  
Qy 279 LQNLSPVDYQOLGNVGTGICMKAGTQALNDICTHRRHSSTSFVYKNGDRAMAKETQFWM 338  
Db 894 -----GTGGAAGNGGDDGTGTTGGDGGAGGTGGTGGTGGGLD----- 931  
Qy 339 DQYVEYFGRPQYQKGPQEVKTDKSWAKALSKPDDGMDTPASMEQFNKAKMIKRPMA 398  
Db 932 ---PRVGG-----SGDGGTGGSGGAAGCGNG---GNAG 960  
Qy 399 DTGNGNLQARGAGGSLGIDAMMAGDAINNMLKGLA 436  
Db 961 AGCNGNGGTCGAGG--IGGTGNGGDAEPGVPPGAGGA 996

RESULT 8  
F70971  
hypotheical glycine-rich protein RV3367 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 18-Feb-2000  
C:Accession: F70971  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: F70971  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-588 <COL>  
A:Cross-references: GB:AL009198; GB:ALL23456; NID:g3242262; PIDN:CAAL15752.1; PID:el102228  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3367  
C:Superfamily: unassigned collagens

Query Match 11.0%; Score 247; DB 2; Length 588;  
Best Local Similarity 30.4%; Pred. No. 2.2e-08;  
Matches 117; Conservative 16; Mismatches 154; Indels 98; Gaps 21;

Qy	52	GGAGG--NNGLLGTSRQNALGGNSALGLGGGNQNDITYNQLAGLLTGMMMMMMWGGGL	109
Db	283	GGAGDANGNPANTSANAGAGGNGAAG-GDGGANGGAGGAGGQAASAGSVGDDGGNGG	341
Qy	110	MGG-----CLGCGLNGLCGCGCLGEGLSNALNDMLGGLNLTGSKGCGNNTTSTFNSPLDQ	165
Db	342	AGGTGTHAGGAG-GAGGAGGGRGWLVG--NGNGNGN---GAAGNGAIGGTCG----	390
Qy	166	ALGNTSTSQNDSTSGTDSTSDSDPMQQLKMFSEIMQSLFGDQDQCTGSSSGGKQPT	225
Db	391	ACGV-PANQCGNSALCT-----OPVCGDGGCGGCGCTCG-----T	424
Qy	226	EGEQNAYKKGYTDALSG-LMGNGLSQLLGNGGLG--GGQGNAGTGLDGGSLGKGLQNL	282
Db	425	GGRGDDGSGGAGASGLWLGNGNG--GNGCTGSGGSGVGGNGTGGDGAG-CGNATSTS	481
Qy	283	SGPDVYQOLGNVAGTGICMKAGIOALNDIGTHRSSRFSFVNKGDRAWAKEIGQFMDIYP	342
Db	482	SIPFDAHG-GNG---GAGGDAG-----HGGTGGDGGDGGHAGTGGRGULL-----	522
Qy	343	EVFGKPOYKGPQGEVKTDDRSWAKLSKPDGDMTASMEQFNKAKGMIKRPMAGDTGN	402
Db	523	-----AGOHANSNGGGGTCGAGTGHG-TPGSGN-----AGGTGT	557
Qy	403	GNLQARGAGGSSLSGIDAMMAGDAIN	427
Db	558	CNADSTNGPGSDG-----LGGDAFN	578

RESULT 9

H70846

hypothetical glycine-rich protein Rv3345c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70846

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: H70846

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1538 <COL>

A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAAL7117.1; PID:g3261517

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv3345c

C:Superfamily: collagen alpha 1(IV) chain

Query Match	10.9%	Score 246;	DB 2;	Length 1538;
Best Local Similarity	27.2%	Pred. No. 7.8e-08;		
Matches 124;	Conservative 33;	Mismatches 168;	Indels 132;	Gaps 21;

Qy	41	GLGASTWQISTGGAGGNNLLGTSRQNALGG-----NSALGLGGGNQNDTYN-----	88
Db	249	GIG-----GTGGNGGAGLFG-----NCGAGGACAAGLPGCAAGLGGDGGDGGNGGTGCGN	298
Qy	89	QLAGLLTGTMMMMMMGGGGLMG-----GCLGGGLCN-CLGGSGG	127
Db	299	GGRGGLLVG-----NGGAGGAGVGVDGGKGGAGDPFSFVNVNAGNGGHHGNPVGVGAGG	354
Qy	128	LGEGLSNALNDMLGCSLNTLGSKGNNTTSTNSPLDQALGINSQNDSTSGTSDTSD	187
Db	355	AG-GLLAGACHAACATPTSCNGCGCGIGATANSPL-QAGCAGCNG-CHGCLVNCGCTGG	411
Qy	188	SSDPMOOLLKMFSEIMQSLFGDGDGTGSSSGGCKQPTGEONAYKKGVTDALSLGMWNG	247
Db	412	AG-----GAGHAGSTGATGTAQPTGGNGT---NGGAGHHGGNGNG	450





A:Accession: F70963  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-778 <COL>  
A:Cross-references: GB:280225; GB:AL123456; NID:g3242265; PIDN:CAB02341.1; PID:e266390;  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv2634c  
C:Superfamily: unassigned collagens

Query Match 10.8%; Score 242; DB 2; Length 778;  
Best Local Similarity 30.1%; Pred. NO. 6.4e-08;  
Matches 91; Conservative 19; Mismatches 110; Indels 82; Gaps 14;

QY 41 GLGASTMTISIGGAG--GNNGLLGTSRQAGLGGNSALGGLG-----GNQNDTVNQLAGLL 94  
DB 353 CNGCNAWPGSCACGCGCTGLACTNGVNPSTANPTGANTGNSGNCQT----- 404  
QY 95 TGMNMMMSMMGGGLMGGGGLG--NGLGGSGGLGEGLSNALNDMLGGSNLTLGSKGG 152  
DB 405 -----GGNG--GPGFAGVGAGVGGGGGLGESLDG--NDGTGG-----KGG 443  
QY 153 NNTTSTNPLDQALGINSSTQNDSTSGTDSDDPMQOALLKMFSEIMQSLFQDQD 212  
DB 444 AGGTAGTGGAGGAGGAGGIGTGTDSAGGV-----ATGGEAGD 481  
QY 213 GTQSSSGGKQPTGE-QNAYKKGVTDAL---SGLMNG-----LSQLLNGGLGG----- 259  
DB 482 CATGVGAGVGCACGCGCGCHTVCVDAPGCGDGGICDGNALGAAGCGCTGCACGNGG 541  
QY 260 -----GOGGNAGTGLDSSSLGKGLQNLSPV---DYQOLGNVAGTGTGCMKAGIQALND 310  
DB 542 RGMGLIGNGAGGAGGTGTGT-GGGGAAGFAGGVGGAGGSELTDGACTAEGTGGTGLGGLG 600  
QY 311 IC 312  
DB 601 VG 602

RESULT 13  
D70835  
hypothetical glycine-rich protein Rv0278c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D70835  
R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: D70835  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-957 <COL>  
A:Cross-references: GB:AL021930; GB:AL123456; NID:g3261524; PIDN:CAA17353.1; PID:g290946  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv0278c  
C:Superfamily: elastin

Query Match 10.7%; Score 241.5; DB 2; Length 957;  
Best Local Similarity 25.8%; Pred. NO. 8.7e-08;  
Matches 117; Conservative 42; Mismatches 149; Indels 145; Gaps 21;

QY 36 SLNTSGLGASTMQISIGGAGGNNLLGTSTRQAGLGG-----NSALGSGGNNQNDTVNQLA 91  
DB 362 TLTAGGICGA-----GAGGNAGLLTSGSGGGAGGFGADGGGGPGGN-----A 407  
QY 92 GLITGMMMMMSMMGGGLMGCGLGGGLGNGLGGSGGL-----CEGLSNAL---N 137  
DB 408 G-----TVFGSGGAGGNGVGGFAGGAGLGGAGGTPLGLGNGGNGGNGGASVATGN 458

QY 138 DMLGSSLNTLGSKGNNTTSTNSPLDQALGINSSTQNDSTSGTDSDDSMQQLK 197  
DB 459 GGIGCTGVLCN--CGCGSGGIGAGKAGVGVGSLLLCLDCFNAPASTPSPLHTLQONVLN 517  
QY 198 MFSEIMQS-----LFGDQDGTGGS-----SSGGKQPTGEGONAYKKGYTDALSG-LKNGG 247  
DB 518 VNNEPFTLTGRPLIGNANGTPTGTGADGGAGWLFNGANGTPTGTGAAGGAGWLFNG 577  
QY 248 LSOLLGNGGLGG-----GOGNAGTGLDSSSLGSKGLONLSGPVDY 288  
DB 578 -----GNGGHGATNTATATGAGGAGGILFTGTGGNGGTG--GIATGAGGIGGAGGAGV 630  
QY 289 QOLGNVAGTG-----ICMKAGIQALNDICTHRSSTRSFVNKGDRAKAIQGFMDQVP 342  
DB 631 SLLIGSGGTGGNGSGISGV-AGIGGAGGRG-----GDAGL----- 664  
QY 343 EYFGKPYQYQPGQEVKTTDDKSWAKALSKPDDGCTFASMEQFNKAKMIKRPMDGTGN 402  
DB 665 -LFG-----AACTGQ-----HCAAGGVDPAGVCCACGN 690  
QY 403 GNLOARCAGSSSLGIDAMMAGDAINNMLKGLG 435  
DB 691 GGLFANGAGGAGGFNA-AGGNGGNGGLFTGG 722

RESULT 14  
H70663  
hypothetical glycine-rich protein Rv1840c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70663  
R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70663  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-515 <COL>  
A:Cross-references: GB:AL123456; NID:g3261678; PIDN:CAB06114.1; PID:gl7812  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv1840c  
C:Superfamily: Phaeolus glycine-rich cell wall protein 1.8

Query Match 10.6%; Score 238.5; DB 2; Length 515;  
Best Local Similarity 27.3%; Pred. NO. 6.6e-08;  
Matches 109; Conservative 23; Mismatches 128; Indels 139; Gaps 15;

QY 52 GCAGGNGGLGTSRQAGLGGNSALGCGGNQNDTVNQLAGLITGMMMSMMGGGGLMG 111  
DB 148 GGAGGSDAGLIGNG-GNGGIGGPGATGLAGG-----AGGVGGLLFGDGGNGGAGGLG 197  
QY 112 GGLGGGLGNGLGGSGGLGEGLSNALNDMLGGSNLTLGSKGNNTTSTNSPL--DQALGI 169  
DB 198 TGPVGATG-CIGGPGGAAGVGLFGHGGAGGAGGLCKAGFAGCAGCTGTGGLLYGNGGNGG 256  
QY 170 NSTSQNDSTSGTDSDDPMQOALLKMFSEIMQSLFQDQDGTGQSSSGGKQPTGEQ 229  
DB 257 NVFSGAADGGAGGDA-----RLINGNGDG-----GSVGAAPTG--- 289  
QY 230 NAYKKGVTDALSCLMGNLSQLLNGGLGQ-----GQGNAGTGLDSSSLGKGLQNLSPV 286  
DB 290 -----ICNG-----GNGGNGWLYGDGSGGSTLQGFSDGGTG----- 322  
QY 287 DYQOLGNVAGTGTGCMKAGIQALNDICTHRSSTRSFVNKGDRAKAIQGFMDQYPEVFG 346  
DB 323 -----CNAGMFGGNGGGSFFDNGDGGTGTGCTLTCNGD----- 358

Search completed: January 22, 2003, 16:28:54  
Job time : 24 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:41 ; Search time 13 Seconds  
(without alignments)  
678.310 Million cell updates/sec

Title: 693\_11\_3

Perfect score: 2248

Sequence: 1 MGFFLSQMPFFLVSTLL.....DAMMAGDAINNALGKLGAA 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2079	92.5	403	10	US-09-086-118-23
2	2079	92.5	403	10	US-09-835-684-3
3	2079	92.5	403	10	US-09-880-371-3
4	2079	92.5	403	10	US-09-879-248-3
5	2079	92.5	403	10	US-09-770-693-3
6	2079	92.5	403	10	US-09-766-348-3
7	718.5	32.0	338	10	US-09-086-118-21
8	718.5	32.0	338	10	US-09-835-684-1
9	718.5	32.0	338	10	US-09-880-371-1
10	718.5	32.0	338	10	US-09-879-248-1
11	718.5	32.0	338	10	US-09-770-693-1
12	718.5	32.0	338	10	US-09-766-348-1
13	211.5	9.4	651	10	US-09-861-597-1
14	198	8.8	606	10	US-09-861-597-6
15	197.5	8.8	344	10	US-09-086-118-27
16	197.5	8.8	344	10	US-09-835-684-11
17	197.5	8.8	344	10	US-09-880-371-11
18	197.5	8.8	344	10	US-09-879-248-15
19	197.5	8.8	344	10	US-09-770-693-7

20	197.5	8.8	344	10	US-09-766-348-7	Sequence 7, Appli
21	197.5	8.8	357	10	US-09-864-761-35807	Sequence 35807, A
22	197	8.8	606	10	US-09-861-597-8	Sequence 8, Appli
23	192.5	8.6	606	10	US-09-861-597-4	Sequence 4, Appli
24	188	8.4	283	10	US-09-864-761-36720	Sequence 36720, A
25	174.5	7.8	440	9	US-10-066-500-106	Sequence 106, App
26	174.5	7.8	440	9	US-10-063-547-52	Sequence 52, Appl
27	174.5	7.8	440	9	US-10-174-590-202	Sequence 202, App
28	174.5	7.8	440	9	US-10-176-758-202	Sequence 202, App
29	174.5	7.8	440	9	US-10-063-616-52	Sequence 52, Appl
30	174.5	7.8	440	9	US-10-175-737-202	Sequence 202, App
31	174.5	7.8	440	12	US-10-006-867-52	Sequence 52, Appl
32	174.5	7.8	440	12	US-10-052-586-202	Sequence 202, App
33	172	7.7	579	9	US-10-108-605-215	Sequence 215, App
34	169	7.5	34	10	US-09-770-693-11	Sequence 11, Appl
35	169	7.5	35	10	US-09-880-371-16	Sequence 16, Appl
36	165	7.3	34	10	US-09-770-693-13	Sequence 13, Appl
37	164.5	7.3	597	10	US-09-793-306-146	Sequence 146, App
38	163.5	7.3	479	10	US-09-918-951-3	Sequence 3, Appli
39	162.5	7.2	201	10	US-09-848-990-22	Sequence 22, Appl
40	162.5	7.2	201	10	US-09-760-364-14	Sequence 14, Appl
41	159.5	7.1	1230	10	US-09-881-752A-150	Sequence 150, App
42	158	7.0	943	9	US-09-996-634-131	Sequence 131, App
43	156	6.9	191	10	US-09-864-761-36985	Sequence 36985, A
44	154.5	6.9	447	10	US-09-835-684-5	Sequence 5, Appli
45	154.5	6.9	447	10	US-09-880-371-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-086-118-23

; Sequence 23, Application US/09086118

; Patent No. US20010011380A1

; GENERAL INFORMATION:

; APPLICANT: Laby, Ronald J.

; APPLICANT: Beer, Steven V.

; APPLICANT: Wei, Zhong-Min

; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR

; TITLE OF INVENTION: FRAGMENTS ELICITING A HYPERSENSITIVE RESPONSE AND USES

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Nixon, Hargrave, Devans & Doyle LLP

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/086,118

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/048,109

; FILING DATE: 30-MAY-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30,727

; REFERENCE/DOCKET NUMBER: 19603/1301

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (716) 263-1304

; TELEFAX: (716) 263-1600

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 403 amino acids

; TYPE: amino acid









```
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Erwinia chrysanthemi
US-09-770-693-1

Query Match      32.0%; Score 718.5; DB 10; Length 338;
Best Local Similarity 42.8%; Pred. No. 3.4e-48;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 47 MOISI-GGAGGNNGLGTSRONAGLGG-NSA-----LGLCCGNQNDTVNQLAGLLTGMWMM 101
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 1 MOITIKAHITGGDLGVSLGAQ--GLKGLNSAASSLSSVDKLSSTIDKLTSAITSMW--- 55
QY 102 SMWGGGLMGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 56 -----FGCALAQLGAS-SKGLCMSNQLGQSGFN-----GAQASNLLSVPK- 96
QY 162 PLDQALGINSTQNDSTSGTSDSSDPMQOILLKMFSEIMQSLFG-----DQDGG 213
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 97 -----SGGDALS-----KMFDAKALDDLGHDTVTVKLTNSNQ 128
QY 214 TQSSSSGGKOPTGEQNAVKKGYTDALSLGMLGNLSQLLGGGGGSGNAGTGLDGSS 273
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 129 LANSMLNASQMTQGNNAFGSGVNNALSSILGNLQSGM-----SGFSQPS 174
QY 274 LGKGLQNLSPVDYQOOLGNNAVGTGKMGAGIOALNDIGTHRHSSSTRSFYNKGDRAK 333
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 175 LGAGGLQCLSGAGAFNOLGNAIGMVGCONAALSALSNSVTHVGNNRHFFVDKEDRGA 234
QY 334 IGQFMDOYPEVFGKPOYKQPGQEVKTTDDKSWAKALSKPDDGDMTPASMEQFNKA 393
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 235 IGQFMDOYPEIFGKPEYQKDGWSSPKTDDKSWAKALSKPDDGDMTPASMEQFNKA 393
QY 394 RPNAGDTGNLQARGAGSSSLGIDAMMAGDAINNMLGKLGAA 437
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGLANA 338

RESULT 12
US-09-766-348-1
; Sequence 1, Application us/09766348
; Patent No. US20020116733A1
; GENERAL INFORMATION:
; APPLICANT: Qiu, Dewen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED RESISTANCE IN PLANTS BY
; FILE REFERENCE: 19603/2986
; CURRENT APPLICATION NUMBER: US/09/766,348
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 08/984,207
; PRIOR FILING DATE: 1997-12-03
; PRIOR APPLICATION NUMBER: 60/033,230
; PRIOR FILING DATE: 1996-12-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Erwinia chrysanthemi
US-09-766-348-1

Query Match      32.0%; Score 718.5; DB 10; Length 338;
Best Local Similarity 42.8%; Pred. No. 3.4e-48;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 47 MOISI-GGAGGNNGLGTSRONAGLGG-NSA-----LGLCCGNQNDTVNQLAGLLTGMWMM 101
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 1 MOITIKAHITGGDLGVSLGAQ--GLKGLNSAASSLSSVDKLSSTIDKLTSAITSMW--- 55
QY 102 SMWGGGLMGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 102 SMWGGGLMGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
QY 102 SMWGGGLMGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 102 SMWGGGLMGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
QY 102 SMWGGGLMGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 102 SMWGGGLMGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
```

```
Db 56 -----FGGALAQLGAS-SKGLCMSNQLGQSGFN-----GAQASNLLSVPK- 96
QY 162 PLDQALGINSTQNDSTSGTSDSSDPMQOILLKMFSEIMQSLFG-----DQDGG 213
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 97 -----SGGDALS-----KMFDAKALDDLGHDTVTVKLTNSNQ 128
QY 214 TQSSSSGGKOPTGEQNAVKKGYTDALSLGMLGNLSQLLGGGGGSGNAGTGLDGSS 273
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 129 LANSMLNASQMTQGNNAFGSGVNNALSSILGNLQSGM-----SGFSQPS 174
QY 274 LGKGLQNLSPVDYQOOLGNNAVGTGKMGAGIOALNDIGTHRHSSSTRSFYNKGDRAK 333
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 175 LGAGGLQCLSGAGAFNOLGNAIGMVGCONAALSALSNSVTHVGNNRHFFVDKEDRGA 234
QY 334 IGQFMDOYPEVFGKPOYKQPGQEVKTTDDKSWAKALSKPDDGDMTPASMEQFNKA 393
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 235 IGQFMDOYPEIFGKPEYQKDGWSSPKTDDKSWAKALSKPDDGDMTPASMEQFNKA 393
QY 394 RPNAGDTGNLQARGAGSSSLGIDAMMAGDAINNMLGKLGAA 437
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGLANA 338

RESULT 13
US-09-861-597-1
; Sequence 1, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-861-597-1

Query Match      9.4%; Score 211.5; DB 10; Length 651;
Best Local Similarity 25.5%; Pred. No. 6.6e-09;
Matches 108; Conservative 21; Mismatches 168; Indels 127; Gaps 17;

QY 41 GLCASTMOISIGGAGGNNGLLGTSRQNAIGLGNLSAL-----GLGGGNQNDTVNQLAGL-- 93
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 104 GGAGAAAAAAGAGAGGGGGGGLGNGAGAGCGCAAAAAAGAGAGCGGVLGSCACRGC 163
QY 94 LTCMMHMMHMMCGGGLMCGGLG--GGLCNGLCGCGGLGSLNSALNDMLGSLNTLGSKG 151
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 164 LGGOGAGAAAAAAGAGAGGGGGGGLGNGAGAGCGGGLGSG-----QGAGRGGLGCG 214
QY 152 GNNTTSTNSPLDQ-ALGINSTQNDDSTSGTSDSSDPMQOILLKMFSEIMQSLFGDC 210
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 215 AGAAAAAAGAGAGAGGGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 256
QY 211 QDGTGSSSSGGKOPTGEQNAVKKGYTDALSLGMLGNLSQLLGGGGGSGNAGTGL 267
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 257 GLSGGAGAGGG-----EG-----AGAAAAAAGAGAGAG-----GYGCGAGAGAGAG 301
QY 268 GLDGSSTLGGKGLAN-----LSCPVDVVOOLCNVAVTIGMKGAGIOALNDIGTHRHSSSTRSF 323
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 302 GSQAGAGGGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342
```





Search completed: January 22, 2003, 16:29:12  
Job time : 14 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:07 ; Search time 18 seconds  
(without alignments)  
714.323 Million cell updates/sec

Title: 693\_ll\_3  
Perfect score: 2248  
Sequence: 1 MGFFLFQMPFFLVSTLL.....DAMAGDAINNALGKLGA 437

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079	92.5	403	2	US-08-200-724A-2
2	2079	92.5	403	2	US-09-030-270A-3
3	2079	92.5	403	4	US-08-851-376A-2
4	2079	92.5	403	4	US-08-984-207-3
5	2079	92.5	403	4	US-09-013-587-3
6	1928	85.8	385	1	US-08-819-539-3
7	1928	85.8	385	5	PCT-US96-08819-3
8	1928	85.8	385	5	PCT-US93-06243-2
9	1913	85.1	338	1	US-08-891-254-1
10	718.5	32.0	338	2	US-08-484-358-2
11	718.5	32.0	338	2	US-08-819-539-1
12	718.5	32.0	338	2	US-09-030-270A-1
13	718.5	32.0	338	3	US-09-118-959-2
14	718.5	32.0	338	4	US-08-984-207-1
15	718.5	32.0	338	4	US-09-013-587-1
16	718.5	32.0	338	5	PCT-US96-08819-1
17	718.5	32.0	338	5	US-09-060-756-727
18	211.5	9.4	651	4	US-08-556-978B-19
19	211.5	9.4	651	4	US-09-247-806-1
20	211.5	9.4	718	1	US-08-425-069-2
21	211.5	9.4	718	1	US-08-317-844B-2
22	211.5	9.4	718	3	US-09-034-177-3
23	211.5	9.4	747	3	US-08-556-978B-63
24	206	9.2	604	4	US-09-060-756-728
25	200	8.9	334	4	US-08-864-038A-3
26	198.5	8.8	738	3	US-09-247-806-6
27	198	8.8	606	4	US-09-247-806-6

28	197.5	8.8	344	1	US-08-891-254-7	Sequence 7, Appli
29	197.5	8.8	344	2	US-08-819-539-7	Sequence 7, Appli
30	197.5	8.8	344	2	US-09-030-270A-7	Sequence 7, Appli
31	197.5	8.8	344	4	US-08-984-207-7	Sequence 7, Appli
32	197.5	8.8	344	4	US-09-013-587-7	Sequence 7, Appli
33	197.5	8.8	344	5	PCT-US96-08819-7	Sequence 23, Appli
34	197	8.8	606	4	US-08-556-978B-23	Sequence 8, Appli
35	197	8.8	606	4	US-09-247-806-8	Sequence 8, Appli
36	196	8.7	1160	3	US-08-808-599A-24	Sequence 24, Appli
37	192.5	8.6	606	4	US-08-556-978B-21	Sequence 21, Appli
38	192.5	8.6	606	4	US-09-247-806-4	Sequence 4, Appli
39	190	8.5	641	4	US-09-249-585A-3	Sequence 3, Appli
40	187.5	8.3	749	1	US-08-317-522A-2	Sequence 2, Appli
41	187.5	8.3	749	1	US-08-439-818A-2	Sequence 2, Appli
42	187.5	8.3	749	2	US-08-751-965-2	Sequence 2, Appli
43	187.5	8.3	749	2	US-08-738-975-2	Sequence 2, Appli
44	187.5	8.3	749	2	US-08-728-626-2	Sequence 2, Appli
45	187.5	8.3	749	3	US-08-808-599A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-200-724A-2  
; Sequence 2, Application US/08200724A  
; Patent No. 5849868  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Bauer, David W.  
; APPLICANT: Beer, Steven V.  
; APPLICANT: Collmer, Alan  
; APPLICANT: He, Sheng-Yang  
; APPLICANT: Laby, Ron J.  
; TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/200,724A  
; FILING DATE: 23-FEB-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/10030  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 403 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-200-724A-2

Query Match 92.5%; Score 2079; DB 2; Length 403;  
Best Local Similarity 100.0%; Pred. No. 5.9e-176;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-851-376A-2

Query Match 92.5%; Score 2079; DB 4; Length 403;  
Best Local Similarity 100.0%; Pred. No. 5.9e-176;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLCASMTQISIGGAGNGLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL 94  
DB 1 MSLNTSGLCASMTQISIGGAGNGLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL 60  
QY 95 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 154  
DB 61 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 120  
QY 155 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 214  
DB 121 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 180  
QY 215 QGSSSGGKQPTGEGQYKGVTDALSGLMGNGLSQLLGGGGLGGGQGNAGTGLDGSLL 274  
DB 181 QGSSSGGKQPTGEGQYKGVTDALSGLMGNGLSQLLGGGGLGGGQGNAGTGLDGSLL 240  
QY 275 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSRFSFVKNKGDRAMAKEI 334  
DB 241 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSRFSFVKNKGDRAMAKEI 300  
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIR 394  
DB 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIR 360  
QY 395 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 437  
DB 361 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 403

RESULT 4  
US-08-984-207-3  
Sequence 3, Application US/08984207  
Patent No. 6235974  
GENERAL INFORMATION:  
APPLICANT: Qiu, Dewen  
APPLICANT: Wei, Zhong-Min  
APPLICANT: Beer, Steven V.  
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED  
TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: P.O. Box 1051, Clinton Square  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/984,207  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/033,230  
FILING DATE: 05-DEC-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/1201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-984-207-3

Query Match 92.5%; Score 2079; DB 4; Length 403;  
Best Local Similarity 100.0%; Pred. No. 5.9e-176;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLCASMTQISIGGAGNGLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL 94  
DB 1 MSLNTSGLCASMTQISIGGAGNGLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL 60  
QY 95 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 154  
DB 61 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 120  
QY 155 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 214  
DB 121 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 180  
QY 215 QGSSSGGKQPTGEGQYKGVTDALSGLMGNGLSQLLGGGGLGGGQGNAGTGLDGSLL 274  
DB 181 QGSSSGGKQPTGEGQYKGVTDALSGLMGNGLSQLLGGGGLGGGQGNAGTGLDGSLL 240  
QY 275 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSRFSFVKNKGDRAMAKEI 334  
DB 241 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSRFSFVKNKGDRAMAKEI 300  
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIR 394  
DB 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIR 360  
QY 395 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 437  
DB 361 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 403

RESULT 5  
US-09-013-587-3  
Sequence 3, Application US/09013587  
Patent No. 6277814  
GENERAL INFORMATION:  
APPLICANT: Qiu, Dewen  
APPLICANT: Wei, Zhong-Min  
APPLICANT: Beer, Steven V.  
TITLE OF INVENTION: ENHANCEMENT OF GROWTH IN PLANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/013,587



STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/819,539  
FILING DATE: 17-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/475,775  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 14603/10050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 385 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-819-539-3

Query Match	85.8%;	Score 1928;	DB 2;	Length 385;				
Best Local Similarity	100.0%;	Pred. No. 1.2e-162;						
Matches 372;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;				
QY	35	MSLNTSGLGASTM	QISIGGAGGNNGL	LGTSRQNLGAGLGGNSALG	GGGQNDTVN	LAGLL	94	
DB	1	MSLNTSGLGASTM	QISIGGAGGNNGL	LGTSRQNLGAGLGGNSALG	GGGQNDTVN	LAGLL	60	
QY	95	TGMMMMMSMMGG	GLGGGLGGGLGG	GLGGGLGGGLGG	GLGGLGSLN	LDMLGSLN	LTGSKGGNN	154
DB	61	TGMMMMMSMMGG	GLGGGLGGGLGG	GLGGGLGGGLGG	GLGSLN	LDMLGSLN	LTGSKGGNN	120
QY	155	TTSTNTSLPDQAL	GINSTSONDST	SGTSDSTSDSG	SDPMQQLL	KMFSETMSQ	SILFGDQDGT	214
DB	121	TTSTNTSLPDQAL	GINSTSONDST	SGTSDSTSDSG	SDPMQQLL	KMFSETMSQ	SILFGDQDGT	180
QY	215	QGSSSGKQKPT	EGQNAYKKGV	TDALSGLMGNGL	SQLLGNGGLGG	GGGQGNAGT	GLDGSSL	274
DB	181	QGSSSGKQKPT	EGQNAYKKGV	TDALSGLMGNGL	SQLLGNGGLGG	GGGQGNAGT	GLDGSSL	240
QY	275	GGKQLNLSPVDY	QOQLGNVGT	GTGIMKAGTQAL	NDIGT	THRHSSTRS	FPVNGDRAMAK	334
DB	241	GGKQLNLSPVDY	QOQLGNVGT	GTGIMKAGTQAL	NDIGT	THRHSSTRS	FPVNGDRAMAK	300
QY	335	GQFMDQYPEV	FGKPYOKG	PGQGVKTTDD	KSAAKALS	KSPDDG	MTPASMEQFNKAK	394
DB	301	GQFMDQYPEV	FGKPYOKG	PGQGVKTTDD	KSAAKALS	KSPDDG	MTPASMEQFNKAK	360
QY	395	PMAGDTGN	GNLIQ	406				
DB	361	PMAGDTGN	GNLIQ	372				

```

RESULT 8
PCI-US96-08819-3
; Sequence 3, Application PC/TUS9608819
; GENERAL INFORMATION:
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/08819  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/475,775  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldmap, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/10051  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 395 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-08819-3

[illegible]

```

RESULT 9
PCT-US93-06243-2
; Sequence 2. Application PC/TUS9306243
; GENERAL INFORMATION:
; APPLICANT: Zhong-Min Wei, David W. Bauer, Steven V.
; APPLICANT: Beer, Alan Collmer, Sheng-Yang He, and Ron J. Laby

```

Query Match 85.1%; Score 1913; DB 5; Length 385;  
Best Local Similarity 99.5%; Pred. No. 2.6e-161;  
Matches 369; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 10  
US-08-891-254-1  
; Sequence 1, Application US/08891254

```

: Patent No. 5776889
: GENERAL INFORMATION:
: APPLICANT: Wei, Zhong-Min
: APPLICANT: Beer, Steven V.
: TITLE OF INVENTION: Hypersensitive Response
: TITLE OF INVENTION: Induced Resistance In Plants
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle
: STREET: Clinton Square, P.O. Box 1051
: CITY: Rochester
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 14603
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/891,254
: FILING DATE: 10-JUL-1997
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/475,775
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 14603/10050
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716) 263-1304
: TELEFAX: (716) 263-1600
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 338 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-891-254-1
:
: Query Match 32.0%; Score 718.5; DB 1; Length 338;
: Best Local Similarity 42.8%; Pred. No. 8.8e-56;
: Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
:
Qy 47 MOISI-GGAGNNGLLGTSRONAGLGG-NSA---LGLGGGNQNDYVNLQAGLLTGMMMM 101
Db 1 MQITIKAHIGDLGVSGVGLGAQ--GLKGLNSAASGLSSVDKLSITIDKLTSAITSM--- 55
Qy 102 SMMGCGGLMGCLGGLGCLNGCGGGLCEGLSNALNDMLGGLSLNTLGSKGNNTTSTNS 161
Db 56 -----FGALAQGLGAS-SKGLGMSNQLGSGFGN-----CAQASNLLSYPK- 96
Qy 162 PLDQALGINTSQNDSTSDTSQDTPMQQLKMFSEIMQSLFG-----DQDQD 213
Db 97 -----SGGDALS-----KMFDAKLDLLGHDTVTKLITNOSNQ 128
Qy 214 TQSSSSGKQTEGEQRYKKGVYTDALSGLMGNGLSQLLNGGLGGCGGAGNAGTGLDGS 273
Db 129 LANSMLNASQMTGNMNAFGSGVNNALSSILCNGLQGSN-----SCFPSQS 174
Qy 274 LGGKCLNLSGPVDYQOLGNNAVGVICMKAGIOALNDIGTHRHSSSTRSFVNKGDRAWAKE 333
Db 175 LGAGGLQGLSAGAFNLGNNAIGVGVQGNALSAISNVSTHVDGNNRHFDKEDRGWAKE 234
Qy 334 ICQFMDQYPEVFGRPYQKQPCQEVKTDKSKWAKALSKPDDDGMTGTPASMQFNKAKGMTK 393
Db 235 ICQFMDQYPEIFGRPEYQKQGVSSPKTDDKSWAKALSKPDDDGMTGASMDKFRQAMGMIK 294
Qy 394 RPMAGDTGNGNLQARGAGGSLGIDAMWAGDAINNALGKLGA 437
Db 295 SAVAGDTGNTNLNRGAGGASLIGDAAVGVGDKIANMSELGLANA 338

```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	32.0%;	Score	718.5;	DB 2;	Length	338;			
Best Local Similarity	42.8%;	Pred. No.	8.8e-56;						
Matches	173;	Conservative	41;	Mismatches	111;	Indels	79;	Gaps	11;
QY	47	MQISI -GGAGNNGLLGTSRONAGLGG -NSA ---LGLGGNQNDVTNQLAGLLTGMMM	101						
Db	1	MQITKAHIGDLGVSLGAQ -GLKGLNSAASLGSVDKLSSTIDKLTLSALTSM	55						
QY	102	SMMGGGGLMGGLGGGLGNGLGGGGLGEGSLGSLNALNMLGGSNLTLGSKGNNNTTSTNS	161						
Db	56	-----FGALAQGLGAS -SKGLGMSNQLGQSGFN ----GAQASNLLSPK -	96						
QY	162	PLDQALGINSTQNDDSTSTGTDSTSDSDPMQQLLKMFSIQLSFG -----DGQDG	213						
Db	97	-----SGDALS -----KMFDRALDQLLGHDTVTKLTNQSNQ	128						
QY	214	TQSSGGKOPTGEONAYKKGVTDALSGLMGNSLGLLNGNGLGGGGNAGTGLDGSS	273						
Db	129	LANSMLNASQWTQGNMAFGSGVNNALSSILGNLGSQM -----SGFSQPS	174						
QY	274	LGCKGLQNLSPVDYQOLGNAGVTGIGMKAGIQALNDIGITHRHSSTRSFVNKGDRMAKE	333						
Db	175	LGAGGLQGLSCAGAFNLGNAIGMGVGNQAALSALSNVSTHVDGNNRHFDVKDRGMAKE	234						
QY	334	IGQFMDQYPEVFGKPYQKPGQCBKVTDDKSWAKALSKPDDDGMTAPSMEOFNKA	393						
Db	235	IGQFMDQYPEIFGKPEYQKDGWSSPKTDDKSWAKALSKPDDDGMTGASMDKFRAMG	294						

```

QY   394 RPMAGDTGNGLQARAGSGSSLGIDMAMGADINNALGKLGAA 437
      :||||| || |::||::|::|::| ::| ::|::|::| ::|
DB    295 SAVAGDGTNTNLNRGAGASIGIDAAVVGVGDKIANNLSLKLANA 338

RESULT 12
US-08-819-539-1
; Sequence 1, Application US/08819539
; Patent No. 5859324
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: Hypersensitive Response
; TITLE OF INVENTION: Induced Resistance In Plants
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,539
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
us-08-819-539-1

Query Match          32.0%; Score 718 5; DB 2; Length 338;
Best Local Similarity 42.8%; Pred. No. 8.8e-56;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

Qy   47 MQIIST-GGAGNGNLLGTSRONAGLGG--NSA---LGLGGNQNDVTYNOLAGLLTGMMMMM 101
      |||: |||: |::|: |::|: |||: |||: |::|: |||: |||: |||:
Db    1 MQITKAHIGDLGVSLGAQ--GLKGLNSRAASSLGSVDKLSLTIDKLTSAITSM--- 55

Qy   102 SMMCGGGLMGCLGGGLGNGLGGSGGLEGISNALNDMLGSLNLTGSKGNNTTSTTNS 161
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    56 -----FGCALAQALGAS-SKGLGMSNLQGSGFN-----CAQAASNLLSVPK- 96

Qy   162 PLDAQALGINSTQNDDTSFTSDTSDSDPQQQLLKWFSEIMQSIFG-----DQDGQ 213
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    97 -----SGGDALS-----KMPDKALDLLGHDTVTKLTNDSNQ 128

Qy   214 TQSGSSGGKQPTEGEONAYKKGVTDALSGLMGNLSOLLGNGGLGGOGGAGTGLDGSS 273
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    129 LANSMILNASMTQGNMNAPFCGVNNALLSILCNGLGQSM-----SGFSQPS 174

Qy   274 LGGKGLONLSPVDYQOGLNAVGTGICMKAGCTQALNDIGTHRHSTRSFVNKGDRAMAKE 333
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    175 LGAGCLOGLSCAGAFNOLIGNAIGMVGOONAALSALSNYSTHVDCNNRHFVKEDRGMAKE 234

```

```

Oy 274 LGGKCLONLSPVDYQOOLGNAVGTICIGKAGIOALNDIGTHRSSSTRSFYNKGDRMAKE 335
    || || || || || || || || || || || || || || || || || || || || || ||
Db 175 LGAAGLGLSGAGAFNLGNLGAIGHGVGQNAALSALSNSVTHVDGNNRHFVDKEDRGMAKE 234
    || || || || || || || || || || || || || || || || || || || || || ||
Qy 334 IGOFMDQYPEVEFGPOYQKGPQEVKTDKSKWAKALSKPDDDGMTSPASMEQFNKAKGMIK 399
    || || || || || || || || || || || || || || || || || || || || || ||
Db 235 IGOFMDQYPEIFGAPETKYKQKDGWSSPKTDDKSWAKALSKPDDDGMTGASMDKFRQAMGMIK 399
    || || || || || || || || || || || || || || || || || || || || || ||
Qy 394 RPMAGDTGNGNLQARGAGGSLGDAMMAGDAIINMALGKLGA 437
    || || || || || || || || || || || || || || || || || || || || || ||
Db 295 SAVACDTCNTNLNRGAGGASLGIDAAVVGDKIANMSLGLANA 338
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 14
US-09-118-959-2
; Sequence 2, Application US/09118959
; Patent No. 6001959
; GENERAL INFORMATION:
; APPLICANT: Bauer, David
; APPLICANT: Collmer, Alan
; TITLE OF INVENTION: Hypersensitive Response Elicitor From
; TITLE OF INVENTION: Erwinia Chrysanthemi
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,959
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/840
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1304
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

US-09-118-959-2          32.0%: Score 718.5; DB 3; Length 338;
Query Match               43.8%; Pred. No. 8.8e-56;
Best Local Similarity     41; Mismatches 11; Indels 79; Gaps 11;
Matches 173; Conservative 41;

Qy 47 MQIST-CGAGCGNGNIGLCHTSRONAGLCG-NSA---LCLGGGNQNDTVNOIACILTCMMMM 101
      |||: ||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MJITIKAHIGDGLVSGLGQAQ--GLKGLNSAASSGVSKLSSTIDKLTSALTSMW---- 55

Qy 102 SNMGCGGLMGGGLNGCGLGNGCGSGLGEGSLNALNDMLGSLNLTLGSKGGNNHTTNTNS 161
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 -----FGALAAQLGAS--SKGLGMNQLQGSGFN-----GAQASNLLSVPK- 96

Qy 162 PLDQALGINSTQNDDSTSGTDSTSDSDPMPMOOLKMFSEIMQSIFG-----DGQDG 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 -----SGGDALS-----KMFDKALDDLGHDTVTKLTNQSNQ 128

```

QY 214 TQSSSSGKQPTGEQONAYKKGVTDALSLMGNLSOLLNGGLGGGGGAGTGLDGSS 273  
Db 129 LANSMLNASQMTQGNMNAFGSGVNNALSSILNGLGQSM-----SGFSQPS 174  
QY 274 LGGKGLQNLGSPVDYQOLGNVAGTGTGIGMKAGTQALNDIGTHRHSSSTRGFVNKGDRAKAKE 333  
Db 175 LGAGGLQGLSGAGAFNOLGNAIGMGVGNALSSALSNVSTHYDGNRRHFVDKEDRGMAKE 234  
QY 334 IGQFMDQYPEVFGKPYQKQGEQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMK 393  
Db 235 IGQFMDQYPEIFGKPYQKQGEQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMK 294  
QY 394 RPMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437  
Db 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGKLANA 338

RESULT 15

US-08-984-207-1  
; Sequence 1, Application US/08984207  
; Patent No. 6235974  
; GENERAL INFORMATION:  
; APPLICANT: Qiu, Dewen  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Beer, Steven V.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED  
; TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/984,207  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/033,230  
; FILING DATE: 05-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1201  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-984-207-1

Query Match 32.0%; Score 718.5; DB 4; Length 338;  
Best Local Similarity 42.8%; Pred. No. 8.8e-56;  
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;  
QY 47 MQIST-GGAGGNGLGTSRQNAIGLG-NSA--LGLGGNQNDVFNQLAGLLTGMHMM 101  
Db 1 MQITKAHIGGDLGVSGILGAQ--GLKGLNSAASSIGSSVDKLSSTIDKLTSLTSM--- 55  
QY 102 SMGGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGSLNTLGSKGNNNTTSTNS 161

Db 56 -----FGGALAQGLGAS-SKGLGMSNQLGOSFGN-----CAQGASNLLSVPK- 96  
QY 162 PLDQALGINSTSONDDSTSGTDTSDSDPMQQLLKMFEIMQSLFG-----DQDQ 213  
Db 97 -----SGGDALS-----KMFDAKALDDLLGHDTFTVTKLTNQS 128  
QY 214 TQSSSSGKQPTGEQONAYKKGVTDALSLMGNLSOLLNGGLGGGGGAGTGLDGSS 273  
Db 129 LANSMLNASQMTQGNMNAFGSGVNNALSSILNGLGQSM-----SGFSQPS 174  
QY 274 LGGKGLQNLGSPVDYQOLGNVAGTGTGIGMKAGTQALNDIGTHRHSSSTRGFVNKGDRAKAKE 333  
Db 175 LGAGGLQGLSGAGAFNOLGNAIGMGVGNALSSALSNVSTHYDGNRRHFVDKEDRGMAKE 234  
QY 334 IGQFMDQYPEVFGKPYQKQGEQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMK 393  
Db 235 IGQFMDQYPEIFGKPYQKQGEQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMK 294  
QY 394 RPMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437  
Db 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGKLANA 338

Search completed: January 22, 2003, 16:26:34  
Job time : 19 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:07 ; Search time 40 Seconds  
(without alignments)

1455.763 Million cell updates/sec

Title: 693\_11\_3

Perfect score: 2248

Sequence: 1 MGFFLFQMPSEFLVSTLLL.....DAMMAGDAINNMAKLGAA 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079	92.5	403	19 AAW75863	Erwinia amylovora
2	2079	92.5	403	19 AAW62455	Erwinia amylovora
3	2079	92.5	403	19 AAW61114	Hypersensitive res
4	2079	92.5	403	20 AAW87639	A hypersensitive r
5	2079	92.5	403	21 AAY71093	Erwinia amylovora
6	2079	92.5	403	21 AAY84854	A hypersensitive r
7	2079	92.5	403	22 AAY84854	Erwinia amylovora
8	2079	92.5	403	23 ABB09225	Erwinia amylovora
9	2079	92.5	403	23 AAE18295	Erwinia amylovora
10	2079	92.5	403	23 AAE16447	E. amylovora hyper

11	1928	85.8	385	18 AAW06598	Hypersensitive res
12	1913	85.1	385	15 AAR45751	Erwinia amylovora
13	718.5	32.0	338	18 AAW06597	Hypersensitive res
14	718.5	32.0	338	20 AAW87638	A hypersensitive r
15	718.5	32.0	338	20 AAW82407	E. chrysanthemi Hr
16	718.5	32.0	338	21 AAY71092	Erwinia chrysanthemi
17	718.5	32.0	338	21 AAY84853	A hypersensitive r
18	718.5	32.0	338	21 AAY55801	E. chrysanthemi hy
19	718.5	32.0	338	22 AAE06709	Erwinia chrysanthemi
20	718.5	32.0	338	23 ABB09224	Erwinia chrysanthemi
21	718.5	32.0	338	23 AAE18294	Erwinia chrysanthemi
22	718.5	32.0	338	23 AAE16446	E. chrysanthemi hy
23	718.5	32.0	340	19 AAW75862	Erwinia chrysanthemi
24	718.5	32.0	340	19 AAW61113	Hypersensitive res
25	713.5	31.7	340	19 AAW62454	Erwinia chrysanthemi
26	226.5	10.1	898	18 AAW31853	Mycobacterium tube
27	224.5	10.0	272	22 ABB70198	Drosophila melanog
28	215.5	9.6	718	12 AAR14308	N.clavipes draglin
29	211.5	9.4	651	20 AAY40097	Spider silk protei
30	211.5	9.4	651	23 AAU11781	Spider natural sil
31	211.5	9.4	718	19 AAW53346	Nephila clavipes s
32	211.5	9.4	718	21 AAY59070	N. clavipes spider
33	206	9.2	604	16 AAR99057	Spider dragline va
34	204	9.1	528	22 AAB82611	Spider recombinant
35	203.5	9.1	318	21 ABB81229	Mycobacterium tube
36	202	9.0	630	23 AAW50042	N. clavipes spidro
37	202	9.0	676	23 AAW50047	N. clavipes spidro
38	202	9.0	989	23 AAW50038	N. clavipes spidro
39	202	9.0	1255	23 AAW50037	N. clavipes spidro
40	202	9.0	1880	23 AAW50039	N. clavipes spidro
41	202	9.0	2280	22 ABB61650	Drosophila melanog
42	201.5	9.0	5002	22 ABB63723	Mycobacterium spec
43	200	8.9	388	20 AAY04999	Arabidopsis thalia
44	199.5	8.9	201	21 AAG31750	New DNA sequence i
45	198.5	8.8	738	19 AAW56163	

#### ALIGNMENTS

RESULT 1  
AAW75863  
ID AAW75863 standard; Protein; 403 AA.  
XX AAW75863:  
XX  
XX 07-DEC-1998 (first entry)  
DT  
DE Erwinia amylovora hypersensitive response elicitor (HRE).  
DE  
XX  
KW Hypersensitive response elicitor; HRE; insect resistance.  
KW biological control; transgenic plant.  
XX  
OS Erwinia amylovora.  
XX  
PN WO9837752-A1.  
XX  
PD 03-SEP-1998.  
XX  
PF 26-FEB-1998; 98WO-US03604.  
XX  
PR 28-FEB-1997; 97US-0039226.  
XX  
(CORR ) CORNELL RES FOUND INC.  
XX  
PI Wei Z, Zitter TA;  
XX  
DR WPI; 1998-495374/42.  
DR N-PSDB; AAV54607.  
XX  
PT use of hypersensitive response elicitor polypeptide - for  
PT application to plants or seeds or transgenic plants or seeds for the  
PT control of insects.

```

XX PS Disclosure; Page 9-10; 75pp; English.
XX PA
XX PI This is the amino acid sequence of a 39 kDa, heat stable
XX CC hypersensitive response elicitor (HRE) of Erwinia amylovora. The
XX CC invention relates to the use of a HRE polypeptide or protein to
XX CC control insects on plants or plants grown from seed treated with HRE.
XX CC Also claimed is a method of insect control for plants that involves:
XX CC (a) providing a transgenic plant or seed transformed with a DNA
XX CC molecule (see AAV54606-09) encoding a HRE polypeptide or protein (see
XX CC AAW75862-67); and (b) growing the transgenic plants or transgenic
XX CC plants produced from the transgenic seeds to control insects. HRE
XX CC prevents direct insect damage to plants by feeding injury. It kills
XX CC insects close to plants, and interferes with insect larval feeding
XX CC on such plants. It also prevents insects from colonising host
XX CC plants and releasing phytotoxins which result in disease damage to
XX CC plants.
XX SQ Sequence 403 AA;

Query Match 92.5%; Score 2079; DB 19; Length 403;
Best Local Similarity 100.0%; Pred. No. 3.5e-155;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLGASTWQISIGGAGGNLLGTSRQNAAGLGGNSALGLGGGQNDVTYNQLAGLL 94
DB 1 MSLNTSGLGASTWQISIGGAGGNLLGTSRQNAAGLGGNSALGLGGGQNDVTYNQLAGLL 60
QY 95 TGMNMMMSMMGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 154
DB 61 TGMNMMMSMMGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
QY 155 TTSTTNSPLDQALGINSTSONDDSTSGTSDTSDSDPMQOLLKMFSEIMOSLFGDGQDGT 214
DB 121 TTSTTNSPLDQALGINSTSONDDSTSGTSDTSDSDPMQOLLKMFSEIMOSLFGDGQDGT 180
QY 215 QGSSSGGKQPTGEONAYKKGVTDALSGLMNGLSQLLGGNGGLGGGGGAGGAGTGLDGSLL 274
DB 181 QGSSSGGKQPTGEONAYKKGVTDALSGLMNGLSQLLGGNGGLGGGGGAGGAGTGLDGSLL 240
QY 275 GKGKQLNLGSPVDYQOLGNVAGTGTGKAGIQAALNDIGTTHRSSTRSFVNKGDRAKAI 334
DB 241 GKGKQLNLGSPVDYQOLGNVAGTGTGKAGIQAALNDIGTTHRSSTRSFVNKGDRAKAI 300
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDDGMTPASMEQFNKAKGMIKR 394
DB 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDDGMTPASMEQFNKAKGMIKR 360
QY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMGALGLGAA 437
DB 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMGALGLGAA 403

RESULT 2
AAW62455
ID AAW62455 standard; Protein; 403 AA.
XX AC AAW62455;
XX DT 09-NOV-1998 (first entry)
XX DE Erwinia amylovora hypersensitive response elicitor (HRE).
XX KW Hypersensitive response elicitor; HRE; growth; transgenic plant.
XX OS Erwinia amylovora.
XX PN WO9832844-A1.
XX PD 30-JUL-1998.
XX * PF 27-JAN-1998; 98WO-US01507.
XX AC AAW61114;

```

```

PR 27-JAN-1997; 97US-0036048.
XX PA (CORR ) CORNELL RES FOUND INC.
XX PI Beer SV, Oiu D, Wei Z;
XX CC WFI; 1998-427940/36.
XX DR N-PSDB; AAV39973.
XX PT Method for enhancing plant growth - comprises use of hypersensitive
XX PT response elicitor polypeptide or protein which may also effect, e.g.
XX PT increase in plant height or earlier germination seed
XX PS Disclosure; Page 15-16; 110pp; English.
XX XX This is the deduced amino acid sequence of the 39 kDa, heat stable
XX CC hypersensitive response elicitor (HRE) of Erwinia amylovora. A
XX CC method of enhancing growth in plants comprises: (a) applying a HRE
XX CC polypeptide or protein in a non-infectious form to a plant or plant
XX CC seed under conditions effective to enhance growth of the plant or
XX CC plants grown from the seed, or (b) providing a transgenic plant or
XX CC plant seed transformed with a DNA molecule encoding a HRE
XX CC polypeptide or protein, and growing the transgenic plant or a plant
XX CC produced from the transgenic seed under conditions effective to
XX CC enhance plant growth. HREs (see AAW62454-59) or nucleic acids
XX CC encoding them (see AAV39972-75) can be used to increase plant growth.
XX CC The HREs may also result in increased plant height and yield, and
XX CC effect early germination and maturation of plant seed and early
XX CC colouration of fruit and plants. E. amylovora HRE can be applied
XX CC to tomato plants to enhance growth without causing disease in that
XX CC species; this bacterium is a pathogen of apple and pear but not
XX CC of tomato.
XX SQ Sequence 403 AA;

Query Match 92.5%; Score 2079; DB 19; Length 403;
Best Local Similarity 100.0%; Pred. No. 3.5e-155;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLGASTWQISIGGAGGNLLGTSRQNAAGLGGNSALGLGGGQNDVTYNQLAGLL 94
DB 1 MSLNTSGLGASTWQISIGGAGGNLLGTSRQNAAGLGGNSALGLGGGQNDVTYNQLAGLL 60
QY 95 TGMNMMMSMMGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 154
DB 61 TGMNMMMSMMGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
QY 155 TTSTTNSPLDQALGINSTSONDDSTSGTSDTSDSDPMQOLLKMFSEIMOSLFGDGQDGT 214
DB 121 TTSTTNSPLDQALGINSTSONDDSTSGTSDTSDSDPMQOLLKMFSEIMOSLFGDGQDGT 180
QY 215 QGSSSGGKQPTGEONAYKKGVTDALSGLMNGLSQLLGGNGGLGGGGGAGGAGTGLDGSLL 274
DB 181 QGSSSGGKQPTGEONAYKKGVTDALSGLMNGLSQLLGGNGGLGGGGGAGGAGTGLDGSLL 240
QY 275 GKGKQLNLGSPVDYQOLGNVAGTGTGKAGIQAALNDIGTTHRSSTRSFVNKGDRAKAI 334
DB 241 GKGKQLNLGSPVDYQOLGNVAGTGTGKAGIQAALNDIGTTHRSSTRSFVNKGDRAKAI 300
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDDGMTPASMEQFNKAKGMIKR 394
DB 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDDGMTPASMEQFNKAKGMIKR 360
QY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMGALGLGAA 437
DB 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMGALGLGAA 403

RESULT 3
AAW61114
ID AAW61114 standard; Protein; 403 AA.
XX AC AAW61114;

```





XX WPI: 2000-303745/26.  
DR N-PSDB; AAAL4938.  
XX  
PT Hypersensitive response elicitor polypeptides useful for imparting  
PT enhanced growth, disease resistance and insect resistance to plants,  
PT especially vegetables and ornamental flowers -  
XX  
XX  
PS Claim 4; Page 8-10; 100pp; English.  
XX  
CC The present sequence represents a hypersensitive response elicitor  
CC polypeptide. The specification describes hypersensitive response  
CC elicitor polypeptide fragments, which do not elicit a hypersensitive  
CC response. Instead, the proteins impart disease resistance to plants,  
CC enhance plant growth, and/or control insects. The polypeptide  
CC fragments may be used to these properties to plants. The plants which  
CC may be treated in this way include vegetables, crops and ornamental  
CC plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,  
CC peanut, corn, potato, sweet potato, bean, pea, chichory, lettuce,  
CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,  
CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,  
CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,  
CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,  
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,  
CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.  
XX  
SQ Sequence 403 AA;  
Query Match 92.5%; Score 2079; DB 21; Length 403;  
Best Local Similarity 100.0%; Pred. No. 3.5e-155;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 35 MSLNTSGLGASTMQIISIGGAGGNNLLGTSRONAGLGCNSALGLGGGNQNDTVNQLAGLL 94  
Db 1 MSLNTSGLGASTMQIISIGGAGGNNLLGTSRONAGLGCNSALGLGGGNQNDTVNQLAGLL 60  
QY 95 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 154  
Db 61 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120  
QY 155 TTSTTNSPLDQALGINSTSONDDSTSGTSDSDPMQQLKMFSEIMQSLFGDGQDGT 214  
Db 121 TTSTTNSPLDQALGINSTSONDDSTSGTSDSDPMQQLKMFSEIMQSLFGDGQDGT 180  
QY 215 QGSSSGGKQPTGEQONAYKGVTDALSGLMGNLSQLLGNGLGGGGGAGTGLDGSLL 274  
Db 181 QGSSSGGKQPTGEQONAYKGVTDALSGLMGNLSQLLGNGLGGGGGAGTGLDGSLL 240  
QY 275 GKGLQNLSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSTSTRSFVNKGDRAAKEI 334  
Db 241 GKGLQNLSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSTSTRSFVNKGDRAAKEI 300  
QY 335 GQFMDQYPEVFGKPOYQKGPQEVKTDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 394  
Db 301 GQFMDQYPEVFGKPOYQKGPQEVKTDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 360  
QY 395 PMAGDTGNGNLQARGAGGSLGIDAMMAGDAINNMALGKLGA 437  
Db 361 PMAGDTGNGNLQARGAGGSLGIDAMMAGDAINNMALGKLGA 403  
RESULT 7  
AAE06710  
ID AAE06710 standard; Protein; 403 AA.  
XX  
AC AAE06710;  
XX  
DT 16-OCT-2001 (first entry)  
XX  
DE Erwinia amylovora hypersensitive response elicitor protein.  
XX  
KW Hypersensitive response elicitor; oomycete; transgenic plant; infection;  
KW gene therapy; crop loss; antifungal.

XX Erwinia amylovora.  
OS  
XX WO200155347-A1.  
PN  
XX 02-AUG-2001.  
PD  
XX 26-JAN-2001; 2001WO-US02579.  
PF  
XX 26-JAN-2000; 2000US-0178565.  
PR  
XX (CORR ) CORNELL RES FOUND INC.  
PA  
XX Beer SV, Bauer DW;  
PI  
XX WPI: 2001-488791/53.  
DR N-PSDB; AAD12806.  
XX  
XX New chimeric gene, useful for controlling plant-pathogenic fungi and  
XX producing oomycete-resistant transgenic plants, comprises first DNA  
XX encoding hypersensitive response elicitor, promoter and regulatory  
XX region -  
PS Claim 9; Page 13-14; 72pp; English.  
XX  
CC The invention relates to a chimeric gene that includes a first DNA  
CC molecule encoding a hypersensitive response elicitor protein or  
CC polypeptide, promoter operably linked 5' to the first DNA molecule  
CC to induce transcription of the first DNA molecule in response to  
CC activation of the promoter by an oomycete and a 3' regulatory region  
CC operably linked to the first DNA molecule. The invention also relates  
CC to a transgenic plant resistant to disease resulting from oomycete  
CC infection, the transgenic plant including the chimeric gene, wherein  
CC the promoter induces transcription of the first DNA molecule in  
CC response to infection of the plant by an oomycete. The chimeric gene  
CC is used in gene therapy. The chimeric gene is useful as an effective  
CC and safe means of controlling plant-pathogenic fungi, particularly  
CC oomycetes, which are responsible for major crop loss and is also useful  
CC for producing transgenic plants of the invention. The present sequence  
CC is Erwinia amylovora hypersensitive response elicitor protein.  
XX  
SQ Sequence 403 AA;  
Query Match 92.5%; Score 2079; DB 22; Length 403;  
Best Local Similarity 100.0%; Pred. No. 3.5e-155;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 35 MSLNTSGLGASTMQIISIGGAGGNNLLGTSRONAGLGCNSALGLGGGNQNDTVNQLAGLL 94  
Db 1 MSLNTSGLGASTMQIISIGGAGGNNLLGTSRONAGLGCNSALGLGGGNQNDTVNQLAGLL 60  
QY 95 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 154  
Db 61 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120  
QY 155 TTSTTNSPLDQALGINSTSONDDSTSGTSDSDPMQQLKMFSEIMQSLFGDGQDGT 214  
Db 121 TTSTTNSPLDQALGINSTSONDDSTSGTSDSDPMQQLKMFSEIMQSLFGDGQDGT 180  
QY 215 QGSSSGGKQPTGEQONAYKGVTDALSGLMGNLSQLLGNGLGGGGGAGTGLDGSLL 274  
Db 181 QGSSSGGKQPTGEQONAYKGVTDALSGLMGNLSQLLGNGLGGGGGAGTGLDGSLL 240  
QY 275 GKGLQNLSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSTSTRSFVNKGDRAAKEI 334  
Db 241 GKGLQNLSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSTSTRSFVNKGDRAAKEI 300  
QY 335 GQFMDQYPEVFGKPOYQKGPQEVKTDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 394  
Db 301 GQFMDQYPEVFGKPOYQKGPQEVKTDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 360  
QY 395 PMAGDTGNGNLQARGAGGSLGIDAMMAGDAINNMALGKLGA 437  
Db 361 PMAGDTGNGNLQARGAGGSLGIDAMMAGDAINNMALGKLGA 403



Db 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403

RESULT 8

ABB09225

ID ABB09225 standard; Protein; 403 AA.

XX ABB09225;

AC ABB09225;

DT 08-JUL-2002 (first entry)

DE Erwinia amylovora hypersensitive response elicitor SRO ID NO:3.

XX Erwinia amylovora hypersensitive response elicitor; fruit; vegetable; plant; desiccation;

KW postharvest disease.

KW postharvest disease.

OS Erwinia amylovora.

OS Erwinia amylovora.

XX WO200180639-A2.

PN 01-NOV-2001.

PD 17-APR-2001; 2001WO-US12468.

PF 19-APR-2000; 2000US-198359P.

PR (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Qiu D, Remick D;

PI WPI; 2002-041357/05.

XX N-PSDB; ABL51710.

DR Inhibiting post harvest disease (caused by Penicillium, Botrytis,

PT Phytophthora, or Erwinia) or desiccation and enhancing the longevity in

PT a fruits or vegetables, using hypersensitive response elicitor proteins

PT or nucleic acids.

XX Example; Page 9-10; 72pp; English.

PS The present invention describes methods for inhibiting post harvest

XX disease or desiccation and enhancing the longevity in a fruits or

CC vegetables, using hypersensitive response elicitor proteins or

CC polypeptides or nucleic acids (I) derived from pathogens (e.g. Erwinia

CC amylovora, E. Stewartii, E. chrysanthemi, E. carotovora, Xanthomonas,

CC Pseudomonas syringae, P. solanacearum, Phytophthora, and Clavibacter).

CC (I) has bactericidal activity, and can be used in gene therapy. The

CC method can be used for inhibiting post harvest disease (caused by

CC Penicillium, Botrytis, Phytophthora, or Erwinia) or desiccation and

CC enhancing the longevity in a fruits or vegetables. The method enables

CC growers, warehouse packers, shippers and suppliers to process, handle

CC and store fruit and vegetables with reduced losses caused by post harvest

CC disease and desiccation, therefore reducing costs to the consumer and

CC improving quality. The present sequence represents a hypersensitive

CC response elicitor protein given in the exemplification of the present

CC invention.

XX Sequence 403 AA;

SQ Query Match 92.5%; Score 2079; DB 23; Length 403;

Best Local Similarity 100.0%; Pred. No. 3.5e-155;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLAGSTMTQISTGGAGNGLLGT SRQNALGNSALGLGGNQNDTVNLAGLL 94

DB 1 MSLNTSGLAGSTMTQISTGGAGNGLLGT SRQNALGNSALGLGGNQNDTVNLAGLL 60

QY 95 TGMAMMMVMCCGGLMGGGLCGGLCGCGGLCEGLSNALNDMLGCSLNTLGSCKGNN 154

DB 61 TGMAMMMVMGGGGGLMGGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 120

QY 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSSDPMOQLLKMFSEIMOSLFGDGODGT 214

Db 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSSDPMOQLLKMFSEIMOSLFGDGODGT 180

QY 215 QGSSSGGKOPTGEQNAYKKGYTDALSGLMGNGLSOLLNGGLGGGQGNAGTGLDGSSL 274

DB 181 QGSSSGGKOPTGEQNAYKKGYTDALSGLMGNGLSOLLNGGLGGGQGNAGTGLDGSSL 240

QY 275 GKGGLONLSGPYDYOOLGNVAGTIGMKKAGIOALNDIGTHRSSSTRSFYNKGDRAKKEI 334

DB 241 GKGGLONLSGPYDYOOLGNVAGTIGMKKAGIOALNDIGTHRSSSTRSFYNKGDRAKKEI 300

QY 335 GGFMPQYPEVFGKPOVKPGQEVKTDKSKAKLSKPDGDDGTMTPASMEQFNKAKGMIKR 304

DB 301 GGFMPQYPEVFGKPOVKPGQEVKTDKSKAKLSKPDGDDGTMTPASMEQFNKAKGMIKR 360

QY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437

DB 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403

RESULT 9

AAE18295

ID AAE18295 standard; Protein; 403 AA.

XX AAE18295;

AC AAE18295;

DT 07-MAY-2002 (first entry)

DE Erwinia amylovora hypersensitive response elicitor (HRE) #1.

XX Hypersensitive response elicitor; HRE; transgenic plant; plant growth;

KW stress tolerance; disease tolerance; modified flower colour;

KW insect resistance; herbicide resistance; male sterility.

XX Erwinia amylovora.

OS Erwinia amylovora.

XX WO2000195724-A2.

PN 20-DEC-2001.

PD 13-JUN-2001; 2001WO-US18955.

PF 15-JUN-2000; 2000US-211585P.

PR (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Derocher J;

PI WPI; 2002-130707/17.

XX N-PSDB; AAD29124.

DR Improving effectiveness of transgenic plants by topical application of

PT a hypersensitive response elicitor protein to the transgenic plant or

PT by incorporating into the plant a transgene encoding the protein -

XX Disclosure; Page 10-11; 86pp; English.

PS The invention relates to methods of improving the effectiveness of

XX transgenic plants which involves either topical application of a

CC hypersensitive response elicitor (HRE) protein to the transgenic plant

CC or incorporating into the transgenic plant a transgene encoding HRE.

CC HRE sequence is used for improving the effectiveness of transgenic

CC plants by maximising the benefit of transgenic traits associated with

CC a deleterious effect on growth, stress tolerance, disease or insect

CC resistance, enhanced growth, herbicide resistance, male sterility,

CC modified flower colour and biochemically modified plant product in

CC the transgenic plants or overcoming the deleterious effects. The

CC present sequence is Erwinia amylovora HRE protein.

XX Sequence 403 AA;

SQ Query Match 92.5%; Score 2079; DB 23; Length 403;

Best Local Similarity 100.0%; Pred. No. 3.5e-155;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLGASTMQISIGGAGGNLLGTSRQAGLGGNSALGLGGGNQNDTVNQLAGLL 94  
DB 1 MSLNTSGLGASTMQISIGGAGGNLLGTSRQAGLGGNSALGLGGGNQNDTVNQLAGLL 60  
QY 95 TGMWMMMSMMGGGLMGGGLGGGLGNGTGGGGGLGEGLSNALNDMLGSLNTLGSKGGNN 154  
DB 61 TGMWMMMSMMGGGLMGGGLGGGLGNGTGGGGGLGEGLSNALNDMLGSLNTLGSKGGNN 120  
QY 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIMQSLFGDGQDGT 214  
DB 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIMQSLFGDGQDGT 180  
QY 215 QGSSSGGKQPTGEGQNAKKGVTDALSLMGNGLSQLLGGNGGCGGAGTGLDGSSL 274  
DB 181 QGSSSGGKQPTGEGQNAKKGVTDALSLMGNGLSQLLGGNGGCGGAGTGLDGSSL 240  
QY 275 GKGGLQNLSPVDYQQLGNVAVGTGIMKAGTQALNDIGTHRSSTRSFVNKGDRAMAKEI 334  
DB 241 GKGGLQNLSPVDYQQLGNVAVGTGIMKAGTQALNDIGTHRSSTRSFVNKGDRAMAKEI 300  
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGTMTPASMEQFNKAKGMIKR 394  
DB 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGTMTPASMEQFNKAKGMIKR 360  
QY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437  
DB 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403  
RESULT 10  
AAE16447  
ID AAE16447 standard; Protein; 403 AA.  
XX  
AC AAE16447;  
XX  
DT 09-APR-2002 (first entry)  
XX  
DE E. amylovora hypersensitive response elicitor protein #1.  
XX  
KW Hypersensitive response elicitor protein; plant growth; fruit coloration;  
KW disease resistance; stress resistance; phytoalexin; insect infection;  
KW plant maturation.  
XX  
OS Erwinia amylovora.  
XX  
FH Key Location/Qualifiers  
FT Domain 32..74  
FT /label= Hypersensitive\_response\_eliciting\_domain\_1  
FT Region 32..57  
FT /label= Acidic\_unit  
FT Region 57..74  
FT /label= Alpha\_helix  
FT Domain 130..180  
FT /label= Hypersensitive\_response\_eliciting\_domain\_2  
FT Region 130..157  
FT /label= Acidic\_unit  
FT Region 157..180  
FT /label= Alpha\_helix  
XX  
PN WO200198501-A2.  
XX  
XX 27-DEC-2001.  
XX  
XX 12-JUN-2001; 2001WO-US18820.  
XX  
XX 16-JUN-2000; 2000US-212211P.  
XX  
XX (EDEN-) EDEN BIOSCIENCE CORP.  
XX  
PI Fan H, Wei Z;  
XX  
DR WPI; 2002-122282/16.

DR N-PSDB; AAD27015.  
XX  
PT New hypersensitive response elicitor proteins comprising spaced apart  
PT domains having an acidic portion linked to an alpha-helix, useful for  
PT imparting disease or stress resistance, controlling insects or  
PT enhancing plant growth -  
XX  
PS Disclosure; Page 9-10; 99pp; English.  
XX  
CC The patent discloses hypersensitive response elicitor proteins and  
CC nucleotides encoding such proteins. Hypersensitive response elicitor  
CC proteins comprise an isolated pair or more of spaced apart domains,  
CC each comprising an acidic portion linked to an alpha-helix and capable  
CC of eliciting a hypersensitive response in plants. Sequences of the  
CC invention are used to impart disease resistance to plants, to enhance  
CC plant growth, to control insects and/or to impart stress resistance  
CC to plants which includes resistance to environmental stresses such as  
CC climate, air pollution, chemical and nutritional stress. The method of  
CC imparting disease resistance has the potential for treating previously  
CC untreatable diseases, treating diseases systemically and avoiding the  
CC use of infectious agents or environmentally harmful materials. Hyper-  
CC sensitive response elicitor sequences are used to enhance plant growth  
CC which encompasses greater yield, increased in quantity of seeds produced,  
CC percentage of seeds germinated, plant size and biomass, bigger fruits,  
CC earlier fruit coloration and plant maturation. They are also used for  
CC insect control which encompasses preventing direct insect damage to  
CC plant by feeding injury, interfering with insect larval feeding on the  
CC plants, preventing insects from colonising host plants and releasing  
CC phytotoxins. Sequences of the invention also prevent subsequent disease  
CC damage to plants resulting from insect infection. The present sequence  
CC is Erwinia amylovora hypersensitive response elicitor protein.  
XX  
SQ Sequence 403 AA;  
Query Match 92.5%; Score 2079; DB 23; Length 403;  
Best Local Similarity 100.0%; Pred. No. 3.5e-155;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 35 MSLNTSGLGASTMQISIGGAGGNLLGTSRQAGLGGNSALGLGGGNQNDTVNQLAGLL 94  
DB 1 MSLNTSGLGASTMQISIGGAGGNLLGTSRQAGLGGNSALGLGGGNQNDTVNQLAGLL 60  
QY 95 TGMWMMMSMMGGGLMGGGLGGGLGNGTGGGGGLGEGLSNALNDMLGSLNTLGSKGGNN 154  
DB 61 TGMWMMMSMMGGGLMGGGLGGGLGNGTGGGGGLGEGLSNALNDMLGSLNTLGSKGGNN 120  
QY 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIMQSLFGDGQDGT 214  
DB 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIMQSLFGDGQDGT 180  
QY 215 QGSSSGGKQPTGEGQNAKKGVTDALSLMGNGLSQLLGGNGGCGGAGTGLDGSSL 274  
DB 181 QGSSSGGKQPTGEGQNAKKGVTDALSLMGNGLSQLLGGNGGCGGAGTGLDGSSL 240  
QY 275 GKGGLQNLSPVDYQQLGNVAVGTGIMKAGTQALNDIGTHRSSTRSFVNKGDRAMAKEI 334  
DB 241 GKGGLQNLSPVDYQQLGNVAVGTGIMKAGTQALNDIGTHRSSTRSFVNKGDRAMAKEI 300  
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGTMTPASMEQFNKAKGMIKR 394  
DB 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGTMTPASMEQFNKAKGMIKR 360  
QY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437  
DB 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403  
RESULT 11  
AAW06598  
ID AAW06598 standard; Protein; 385 AA.  
XX  
AC AAW06598;  
XX

```

DF 30-MAR-1997 (first entry)
xx Hypersensitive response elicitor protein.
xx
xx Hypersensitive response; elicitor: Erwinia amylovora; plant;
xx disease-resistance; Escherichia coli; infiltration; virus;
xx bacterium; fungus; pathogen; biological control agent.
xx
xx Erwinia amylovora.
xx
xx W09639802-A1.
xx
xx 19-DEC-1996.
xx
xx 05-JUN-1996; 96WO-US08819.
xx
xx 07-JUN-1995; 95US-0475775.
xx
xx (CORR ) CORNELL RES FOUND INC.
xx
xx Beer SV, Wei Z;
xx
xx WPI: 1997-051614/05.
xx
xx N-PSDB; AAR49314.
xx
xx Imparting pathogen resistance to plants - with hypersensitive
xx response elicitor polypeptide or protein
xx
xx Claim 7; Page 46-47; 69pp; English.
xx
xx This sequence represents a hypersensitive response elicitor from
xx Erwinia amylovora, with a mol.wt. of 37,000. The elicitor has a pI
xx of 4.3, thermostability at 100 deg C for at least 10 min, and
xx contains no cysteine. The elicitor may be used in a new method for
xx imparting pathogen resistance to plants, by application of the
xx elicitor in a non-infectious form to plant cells, by spraying,
xx injection, leaf abrasion, or plant infection with recombinant
xx bacteria (non-infectious to the host plant, e.g. Escherichia coli)
xx expressing the elicitor as a biological control agent, to allow
xx recombinant protein infiltration into the plant. The method
xx confers virus, bacterium or fungus disease-resistance on crops and
xx ornamental plants.
xx
xx Sequence 385 AA:
xx
xx Query Match 85.8%; Score 1928; DB 18; Length 385;
xx Best Local Similarity 100.0%; Pred. NO. 2.4e-143;
xx Matches 372; Conservative 0; Mismatches 0; Indels 0; Caps 0;
xx
xx QY 35 MSLNTSGLGASTMQISIGGAGGNGLLGTSRONAGLGGNSALGLGGGNDTVNQLAGLL 94
xx DB 1 MSLNTSGLGASTMQISIGGAGGNGLLGTSRONAGLGGNSALGLGGGNDTVNQLAGLL 60
xx
xx QY 95 TGMAMMMNMGGGLMGGGLGGGLGNGLCGGGGLGEGLSNALNDMLGGSINTLGSKGNN 154
xx DB 61 TGMAMMMNMGGGLMGGGLGGGLGNGLCGGGGLGEGLSNALNDMLGGSINTLGSKGNN 120
xx
xx QY 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIHQSLFQDGDGT 214
xx DB 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIHQSLFQDGDGT 180
xx
xx QY 215 QCSSSGGKOPTGEQNAKKYKGVTDALSLMGNGLSOLLNGGLGGCGQGNAGTGLDGSLL 274
xx DB 181 QCSSSGGKOPTGEQNAKKYKGVTDALSLMGNGLSOLLNGGLGGCGQGNAGTGLDGSLL 240
xx
xx QY 275 GSKGLONLGGPVDOVLGNVGTGTCMRAGTQALNDIGTHRSSSTRSFVNGKDRAMAKET 334
xx DB 241 GSKGLONLGGPVDOVLGNVGTGIGMKRAGIQALNDIGTHRSSSTRSFVNGKDRAMAKET 300
xx
xx QY 335 GQFMDQYFVFGKPYQKGPQGVKRTDDKSWAKALS KPDGDTGWT FASMDQFNKAKGMIR 394
xx DB 301 GQFMDQYFVFGKPYQKGPQGVKRTDDKSWAKALS KPDGDTGWT FASMDQFNKAKGMIR 360

```

```

QY 395 PMAGDTGNGNLQ 406
DB 361 PMAGDTGNGNLQ 372
xx
xx RESULT 12
xx AAR45751
xx ID AAR45751 standard; Protein: 385 AA.
xx
xx AAR45751;
xx
xx 11-JUL-1994 (first entry)
xx
xx Erwinia amylovora harpin.
xx
xx Harpin: hypersensitive response elicitor; HR-elicitor; fire blight;
xx Rosaceae; apple; pear; phytopathogenic bacteria; defence reaction;
xx hypersensitive reaction and pathogenicity; hrpN; gene cluster; ds.
xx
xx Erwinia amylovora.
xx
xx Key Location/Qualifiers
xx Misc-difference 372 /note= "corresponds to CAG codon"
xx FT MISC-difference 373 /note= "corresponds to CAC codon"
xx
xx W09401546-A.
xx
xx 20-JAN-1994.
xx
xx 30-JUN-1993; 93WO-US06243.
xx
xx 01-JUL-1992; 92US-0907935.
xx
xx (CORR ) CORNELL RES FOUND INC.
xx
xx Bauer DW, Beer SV, Collmer A, He S, Laby R, Wei Z;
xx
xx WPI: 1994-035054/04.
xx
xx N-PSDB; AAO55751.
xx
xx Hypersensitive response elicitor protein derived from Erwinia
xx amylovora - and DNA encoding it, useful for developing harpin
xx inhibitors to prevent e.g. fire blight of fruit
xx
xx Claim 6; Page 27-28; 47pp; English.
xx
xx The hrpN gene was isolated from E.amylovora using a 48-fold
xx degenerate oligonucleotide probe corresponding to amino acids 9-15
xx at the N-terminus of harpin. The 44kD protein encoded by the hrpN
xx gene is a hypersensitive response elicitor protein. The harpin is
xx thought to be an archetype for HR elicitors from phytopathogenic
xx bacteria.
xx
xx Sequence 385 AA;
xx
xx Query Match 85.1%; Score 1913; DB 15; Length 385;
xx Best Local Similarity 99.5%; Pred. NO. 3.6e-142;
xx Matches 369; Conservative 1; Mismatches 1; Indels 0; Caps 0;
xx
xx QY 35 MSLNTSGLGASTMQISIGGAGGNGLLGTSRONAGLGGNSALGLGGGNDTVNQLAGLL 94
xx DB 1 MSLNTSGLGASTMQISIGGAGGNGLLGTSRONAGLGGNSALGLGGGNDTVNQLAGLL 60
xx
xx QY 95 TGMAMMMNMGGGLMGGGLGGGLGNGLCGGGGLGEGLSNALNDMLGGSINTLGSKGNN 154
xx DB 61 TGMAMMMNMGGGLMGGGLGGGLGNGLCGGGGLGEGLSNALNDMLGGSINTLGSKGNN 120
xx
xx 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIHQSLFQDGDGT 214
xx DB 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLKMFSEIHQSLFQDGDGT 180

```

QY 215 QGSSCGKQPTGEQONAYKKVTDALSGLMGNLSQLLNGGLGGGQGNAGTGLDGSSL 274  
 DB 181 QGSSGGKQPTGEQONAYKKVTDALSGLMGNLSQLLNGGLGGGQGNAGTGLDGSSL 240  
 QY 275 GKGGLNLSPVDYQOGLGNVGTGIGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 334  
 DB 241 GKGGLNLSPVDYQOGLGNVGTGIGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 300  
 QY 335 GQFMDQYPEVFGPKQYQKGPQEQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIKR 394  
 DB 301 GQFMDQYPEVFGPKQYQKGPQEQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIKR 360  
 QY 395 PMAGDTGNGNL 405  
 DB 361 PMAGDTGNGNL 371

RESULT 13  
 AAW06597  
 ID AAW06597 standard; Protein; 338 AA.  
 AC AAW06597;  
 DT 30-MAR-1997 (first entry)  
 DE Hypersensitive response elicitor protein.  
 KW Hypersensitive response; elicitor; Erwinia chrysanthemi; plant;  
 KW disease-resistance; Escherichia coli; infiltration; virus;  
 KW bacterium; fungus; pathogen; biological control agent.  
 OS Erwinia chrysanthemi.  
 PN WO9639802-A1.  
 PD 19-DEC-1996.  
 PF 05-JUN-1996; 96WO-US08819.  
 PR 07-JUN-1995; 95US-0475775.  
 PA (CORR ) CORNELL RES FOUND INC.  
 PI Beer SV, Wei Z;  
 DR WPI: 1997-051614/05.  
 DR N-PSDB; AAY49313.  
 PT Imparting pathogen resistance to plants - with hypersensitive  
 PT response elicitor polypeptide or protein  
 PS Claim 4; Page 44; 69pp; English.  
 CC This sequence represents a hypersensitive response elicitor from  
 CC Erwinia chrysanthemi, with a mol.wt. of 34,000. The elicitor is  
 CC thermostable, has a glycine content of over 16%, and has no cysteine.  
 CC The elicitor may be used in a new method for imparting pathogen  
 CC resistance to plants, by application of the elicitor in a non-  
 CC infectious form to plant cells, by spraying, injection, leaf  
 CC abrasion, or plant infection with recombinant bacteria (non-  
 CC infectious to the host plant, e.g. Escherichia coli) expressing the  
 CC elicitor as a biological control agent, to allow recombinant  
 CC protein infiltration into the plant. The method confers virus,  
 CC bacterium or fungus disease-resistance on crops and ornamental  
 CC plants.  
 CC Sequence 338 AA;  
 SQ

Query Match 32.0%; Score 718.5; DB 18; Length 338;  
 Best Local Similarity 42.8%; Pred. No. 1.9e-48;  
 Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 47 MQTISI-GGAGGNNLLGTSRQNLGG-NSA---LGLGGGNDNTVNQLAGLLTGMMMM 101

DB 1 MQITIKAHITGDLGVSGLGAQ--GLKGLNSAASSLGSVDKLSSTIDKLTSAITSMW--- 55  
 QY 102 SMWGGGLMGGLGGLGNGLGGSGGLGELSNALNDMLGGLSLNTLGSKGGNNTTSTNS 161  
 DB 56 -----FGGALAAGLGAS-SKGLGMSNQLGQSFGN-----GAQASNLLSVPK- 96  
 QY 162 PLDQALINSTQNDSDSTSDSDPMQQLLKMFSIMQSLFG-----DQDQG 213  
 DB 97 -----SGGDALS-----KMFDAKLDLLGHDTVTVLTNQSNQ 128  
 QY 214 TQSSSGGKQPTGEQONAYKKVTDALSGLMGNLSQLLNGGLGGGQGNAGTGLDGSS 273  
 DB 129 LANSMLNASQMTQGNMNAFGSGVNNALSSILGNLQOSM-----SGFSQPS 174  
 QY 274 LGGKGLNLSPVDYQOGLGNVGTGIGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 333  
 DB 175 LGAGGLQGLSGAGAFNQLGNALGMGVQNAALSAJSNVSTHVDGNRHFVDKEDQMAKE 234  
 QY 334 IGQFMDQYPEVFGPKQYQKGPQEQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIK 393  
 DB 235 IGQFMDQYPEVFGPKQYQKGDHSSPKTDDKSWAKALSKPDDGDMTGASMDKFRQAMGMIK 294  
 QY 394 RPMAGDTGNGNLQARGAGSSGLGIDAMMAGDAIINNMLGKLGAA 437  
 DB 295 SAVAGDTGNTNLNRGAGGASLGIDAAYVGDKIANNLSLGLKLANA 338

RESULT 14  
 AAW87638  
 ID AAW87638 standard; Protein; 338 AA.  
 AC AAW87638;  
 DT 09-MAR-1999 (first entry)  
 DE A hypersensitive response elicitor protein.  
 KW Hypersensitive response elicitor protein; hairpin protein;  
 KW disease resistance; seed quality; insect control; corn borer;  
 KW Lepidoptera larvae; transgenic plant.  
 OS Erwinia chrysanthemi.  
 PN WO9854214-A2.  
 PD 03-DEC-1998.  
 PF 28-MAY-1998; 98WO-US10874.  
 PR 30-MAY-1997; 97US-0048109.  
 PA (CORR ) CORNELL RES FOUND INC.  
 PA (EDEN-) EDEN BIOSCIENCE CORP.  
 PI Beer SV, Laby RJ, Wei Z;  
 DR WPI: 1999-070210/06.  
 DR N-PSDB; AAV83988.  
 PT New fragments of an Erwinia hypersensitive response elicitor protein  
 PT and related DNA - used to impart disease resistance to plants, to  
 PT increase their growth and to control insects  
 PS Disclosure; Page 7-8; 94pp; English.  
 CC The present sequence represents a hypersensitive response elicitor  
 CC protein (also called hairpin protein) that is able to elicit a  
 CC hypersensitive response in plants. The specification also describes  
 CC hypersensitive response elicitors from other pathogenic organisms.  
 CC The protein, in non-infectious form, is applied to plants to impart  
 CC disease resistance (to a wide range of viral, bacterial and fungal  
 CC pathogens), to improve growth (yield, quantity and quality of seeds,

CC to provide earlier germination etc.) and to control insects (e.g. corn  
 CC borers, Lepidoptera larvae etc.) The same results are provided by  
 CC transgenic plants expressing the protein.

XX  
 SQ Sequence 338 AA:  
 Query Match 32.0%; Score 718.5; DB 20; Leng-h 338;  
 Best Local Similarity 42.8%; Pred. NO. 1.9e-48;  
 Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;  
 QY 47 MOISI-GGAGGNNLLCTSRONAGLGG-NSA---LGLGGGNQNDTVNOLAGLLTGMMM 101  
 DB 1 MQITIKAHIGDGLVSGLGAQ--GLKGLNSAASSLGSSVDKLSSTIDKLTSAITSM--- 55  
 QY 102 SMGGGGLMGGGLGCGLNGLGGCGGCGEGLSNALNDMLGSLNTLGSKGGNNTTSTTNS 161  
 DB 56 -----FGGALAOGLGAS--SKGLGMSNQLGQSFGN-----GAQGASNLLSVPK- 96  
 QY 162 PLDQALGINSTSQNDSTSGTSDTSDSDPQQLLKMFSEIMQSLFG-----DQDQG 213  
 DB 97 -----SGGDALS-----KMFDAKALDDLLGHDTVTKLTNOSNQ 128  
 QY 214 TOGSSSGGKQPTGEONAYKKGYTDALSGLMGNGLSQLLGGNGGGGGNAGTGLDGSS 273  
 DB 129 LANSMLNASQMTQGNMNAFGSVNNALSSILNGLGQSM-----SGFSQPS 174  
 QY 274 LGCKGLONLSCPVDYOOLGNVCTGICMKACTQALNDICTHRHSSTFRSVNKGDRAMAKE 333  
 DB 175 LGAGGLGSLGAGAFNOLGNAIGMGVGONAAJLSJNSVTHVDGNNRHVFYDKEDRGMAKE 234  
 QY 334 IGQFMDQYPEVFGKPYQKQGVKTDKSWAKALSKPDDDGMTGASMEQFNKAKGMK 393  
 DB 235 IGQFMDQYPEIFGKPEYOKDCHWSSPKTDDKSWAKALSKPDDDGMTGASMDKFRQAMGMK 294  
 QY 394 RPMAGDTGNGNLQARGAGSSGLGIDAMMAGDAINNMALGKLGA 437  
 DB 295 SAVAGDTGNTNLRGAGGASLGIDAAVVGDKIANNSLGKLANA 338

## RESULT 15

AAW82407

ID AAW82407 standard; Protein; 338 AA.

XX AC AAW82407;

XX DT 23-FEB-1999 (first entry)

XX DE E. chrysanthemi HrpN-ECH protein.

XX KW Hypersensitive response elicitor protein; hrpN-Ech; pathogen resistance;

XX KW plant; transformation; pathogen-inducible promoter.

XX OS Erwinia chrysanthemi.

XX PN U95850015-A.

XX PD 15-DEC-1998.

XX PF 07-JUN-1995; 95US-0484358.

XX PR 07-JUN-1995; 95US-0484358.

XX PA (CORR ) CORNELL RES FOUND INC.

XX PI Bauer D, Collmer A;

XX WI: 1999-069852/06.

XX DR N-PSDB; AAV73494, AAV73507.

XX PT DNA encoding Erwinia chrysanthemi hypersensitive response protein

XX PT hrpN - useful for imparting pathogen resistance to plants

XX PS Claim 2; Column 29-30; 27pp; English.

XX This sequence represents a novel Erwinia chrysanthemi protein, hrpN-Ech,  
 CC that elicits a hypersensitive response in plants. The encoding DNA can be  
 CC used for imparting pathogen resistance to plants, by transforming a  
 CC plant with a vector containing the DNA and a pathogen-inducible promoter.

XX  
 SQ Sequence 338 AA:

Query Match 32.0%; Score 718.5; DB 20; Length 338;  
 Best Local Similarity 42.8%; Pred. NO. 1.9e-48;  
 Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;  
 QY 47 MOISI-GGAGGNNLLCTSRONAGLGG-NSA---LGLGGGNQNDTVNOLAGLLTGMMM 101  
 DB 1 MQITIKAHIGDGLVSGLGAQ--GLKGLNSAASSLGSSVDKLSSTIDKLTSAITSM--- 55  
 QY 102 SMGGGGLMGGGLGCGLNGLGGCGGCGEGLSNALNDMLGSLNTLGSKGGNNTTSTTNS 161  
 DB 56 -----FGGALAOGLGAS--SKGLGMSNQLGQSFGN-----GAQGASNLLSVPK- 96  
 QY 162 PLDQALGINSTSQNDSTSGTSDTSDSDPQQLLKMFSEIMQSLFG-----DQDQG 213  
 DB 97 -----SGGDALS-----KMFDAKALDDLLGHDTVTKLTNOSNQ 128  
 QY 214 TOGSSSGGKQPTGEONAYKKGYTDALSGLMGNGLSQLLGGNGGGGGNAGTGLDGSS 273  
 DB 129 LANSMLNASQMTQGNMNAFGSVNNALSSILNGLGQSM-----SGFSQPS 174  
 QY 274 LGCKGLONLSCPVDYOOLGNVCTGICMKACTQALNDICTHRHSSTFRSVNKGDRAMAKE 333  
 DB 175 LGAGGLGSLGAGAFNOLGNAIGMGVGONAAJLSJNSVTHVDGNNRHVFYDKEDRGMAKE 234  
 QY 334 IGQFMDQYPEVFGKPYQKQGVKTDKSWAKALSKPDDDGMTGASMEQFNKAKGMK 393  
 DB 235 IGQFMDQYPEIFGKPEYOKDCHWSSPKTDDKSWAKALSKPDDDGMTGASMDKFRQAMGMK 294  
 QY 394 RPMAGDTGNGNLQARGAGSSGLGIDAMMAGDAINNMALGKLGA 437  
 DB 295 SAVAGDTGNTNLRGAGGASLGIDAAVVGDKIANNSLGKLANA 338

Search completed: January 22, 2003, 16:28:24

Job time : 42 secs